## (12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION EN MATIÈRE DE BREVETS (PCT)

# (19) Organisation Mondiale de la Propriété Intellectuelle

Bureau international



# Rec'd PCT/PTO 13 MAR 2005

(43) Date de la publication internationale 1 avril 2004 (01.04.2004)

**PCT** 

# (10) Numéro de publication internationale WO 2004/027068 A2

- (51) Classification internationale des brevets<sup>7</sup>: C12N 15/67
- (21) Numéro de la demande internationale :

PCT/FR2003/002763

(22) Date de dépôt international:

19 septembre 2003 (19.09.2003)

(25) Langue de dépôt :

français

(26) Langue de publication :

français

(30) Données relatives à la priorité :
02/11676 20 septembre 2002 (20.09.2002) FR

- (71) Déposant (pour tous les États désignés sauf US): CENTRE NATIONAL DE LA RECHERCHE SCIEN-TIFIQUE [FR/FR]; 3, rue Michel Ange, F-75794 Paris Cedex 16 (FR).
- (72) Inventeurs; et
- (75) Inventeurs/Déposants (pour US seulement): FALSON, Pierre [FR/FR]; 5, avenue Valioud, F-69110 Sainte Foy les Lyon (FR). PENIN, François [FR/FR]; 20, avenue des Platanes, F-69150 Decines Charpieu (FR). MONTIGNY, Cédric [FR/FR]; Résidence La Cilof, Bât. K., Route de Chateaufort, F-91190 Gif sur Yvette (FR).

- (74) Mandataire: AUDIER, Philippe; c/o Brevatome, 3, rue du Docteur Lancereaux, F-75008 Paris (FR).
- (81) États désignés (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) États désignés (régional): brevet ARIPO (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), brevet eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

#### Publiée:

 sans rapport de recherche internationale, sera republiée dès réception de ce rapport

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

- (54) Title: SYSTEMS FOR EXPRESSING TOXIC PROTEINS, VECTORS AND METHOD OF PRODUCING TOXIC PROTEINS
- (54) Titre : SYSTEMES D'EXPRESSION DE PROTEINES TOXIQUES, VECTEURS ET PROCEDE DE FABRICATION DE PROTEINES TOXIQUES
- (57) Abstract: The invention relates to a system for expressing toxic proteins, an expression vector comprising the aforementioned system, a prokaryote cell transformed by the system and a method of synthesising a toxic protein using said expression system. Said expression system is characterised in that it comprises the following successive sequences in direction 5' 3', namely a nucleotide sequence coding for dipeptide Asp-Pro and a nucleotide sequence coding for a toxic protein. In a particularly preferable mode, the expression system also comprises, upstream of the Asp-Pro sequence, a nucleotide sequence coding for a soluble protein. The inventive expression system can be used to construct an expression vector which can be used to transform a prokaryote cell such as E. coli, e.g. in a toxic protein synthesis method.
  - (57) Abrégé: La présente invention se rapporte à un système d'expression de protéines toxiques, à un vecteur d'expression comprenant ce système, à une cellule procaryote transformée par ce système, ainsi qu'à un procédé de synthèse d'une protéine toxique utilisant ce système d'expression. Le système d'expression de l'invention se caractérise en ce qu'il comprend successivement, dans le sens 5'-3', une séquence nucléotidique codant pour le dipeptide Asp-Pro et une séquence nucléotidique codant pour une protéine toxique. Selon un mode particulièrement préféré de l'invention, le système d'expression comprend en outre, en amont de la séquence Asp-Pro, une séquence nucléotidique codant pour une protéine soluble. Le système d'expression de l'invention permet de construire un vecteur d'expression utile pour la transformation d'une cellule procaryote telle que E. coli, par exemple dans un procédé de synthèse de la protéine toxique.



# SYSTEMES D'EXPRESSION DE PROTEINES TOXIQUES, VECTEURS ET PROCEDE DE FABRICATION DE PROTEINES TOXIQUES

#### DESCRIPTION

## 5 DOMAINE TECHNIQUE

La présente invention se rapporte à des systèmes d'expression de protéines toxiques, à des vecteurs d'expression comprenant un de ces systèmes, à des cellules procaryotes transformées par ces systèmes, ainsi qu'à un procédé de synthèse d'une protéine toxique utilisant ces systèmes d'expression.

Elle permet par exemple la surproduction dans une cellule procaryote, par exemple Escherichia coli (E.coli), de protéines ou peptides hydrophobes toxiques, par exemple la surproduction de domaines transmembranaires de protéines d'enveloppe de virus.

Elle trouve de nombreuses applications notamment dans la recherche concernant les mécanismes d'infections virales et dans recherche et la mise au point de nouveaux principes actifs pour lutter contre les infections virales.

Dans la description qui suit, les références entre crochets [] renvoient à la liste de références annexée.

25

30

20

10

15

# Etat de la technique

La détermination de la structure tridimensionnelle (3D) est une étape décisive dans la compréhension structurale et fonctionnelle des protéines.

Γ/FR2003/002763

5

10

15

20

25

De très grands efforts et moyens ont été, et sont, mis en oeuvre pour parvenir à ce but, et se sont amplifiés avec l'accumulation des données apportées par les programmes de séquençage des génomes [1].

Les deux principales techniques permettant l'établissement de ces structures des protéines sont la diffraction des rayons X, réalisée à partir de protéines cristallisées, et la résonance magnétique nucléaire (RMN) effectuée à partir de protéines en solution. La RMN, bien adaptée à l'étude de protéines de masse moléculaire inférieure à 20 kDa, nécessite cependant, comme la diffraction des RX, la production de grandes quantités de matériel. Elle implique en outre dans la plupart des cas de préparer du matériel enrichi en <sup>15</sup>N et/ou <sup>13</sup>C.

Dans ce contexte, la bactérie est un moyen de production largement utilisé par la communauté scientifique [2]. La surexpression de protéines dans la bactérie ne s'effectue cependant pas sans problèmes. En effet, elle donne lieu à trois cas de figures :

- Le premier cas, idéal, est celui où la protéine est surproduite sous une forme correctement repliée dans l'espace lors de sa synthèse in vivo. Ce n'est pas une situation rare, mais elle n'est pas fréquente non plus. Elle concerne essentiellement les protéines solubles et de taille réduite, c'est à dire d'environ 20 à 50 kDa.
- Le deuxième cas, le plus courant, est celui où la protéine est surproduite et agrégée sous la forme
   de corps d'inclusion. Cela concerne les protéines polytopiques et/ou de grande taille. Dans ce cas, la

15

20

25

30

cinétique de repliement de la protéine est nettement plus lente que sa vitesse de biosynthèse. Cela favorise l'exposition au solvant aqueux des régions hydrophobes de la protéine qui sont normalement enfouies au coeur de celle-ci, et génère des interactions non spécifiques qui conduisent à la formation d'agrégats insolubles. Suivant le degré de désordre de ce repliement, les corps d'inclusions peuvent être solubilisés/dépliés dans des conditions non natives, avec de l'urée ou de la guanidine. La protéine solubilisée est ensuite soumise à différents traitements comme la dialyse ou la dilution pour favoriser, dans certains cas avec succès, un repliement 3D natif.

- Le troisième cas est celui où l'expression engendre une toxicité d'un degré variable. Celui-ci va de l'absence de produit d'expression si la bactérie parvient à s'adapter, à la mort de celle-ci si le produit est trop toxique. C'est un cas qui se présente assez fréquemment et le plus souvent avec des protéines ou des domaines de protéines membranaires, comme par exemple ceux des protéines d'enveloppe du virus de l'hépatite C [5] ou du virus d'immunodéficience humain [6].

problème toxicité concerne Le de essentiellement l'expression de protéines membranaires, c'est à dire de protéines présentant un domaine hydrophobe. Or ces protéines présentent un intérêt Elles sont d'une part relativement grandissant. l'établissement des différents puisque nombreuses génomes confirme qu'elles représentent environ 30% des protéines potentiellement codées par ces génomes [7].

10

15

20

25

30

Elles constituent d'autre part 70% des cibles thérapeutiques et leur altération est à l'origine de nombreuses maladies génétiques [8].

Il est donc essentiel de mettre au point des méthodes facilitant ou permettant l'expression de telles protéines ou de leur partie membranaire.

Des efforts ont été faits dans ce sens avec par exemple la mise au point de souches bactériennes, soit tolérant mieux l'expression de protéines membranaires [9, 10], soit ayant une régulation plus stricte du mécanisme d'expression comme dans le cas de la souche E. coli BL21(DE3)pLysS développée par Stratagene. Cependant, ces améliorations ne permettent pas de supprimer le phénomène de toxicité dans tous les cas, en particulier lors de l'expression de peptides hydrophobes correspondant à des ancres membranaires.

enjeux médicaux majeurs actuels L'un des concerne le traitement de l'hépatite C. Celle-ci est provoquée par le virus de l'hépatite C (VHC) de la famille des flaviviridae et qui infecte spécifiquement les cellules hépatiques [11]. Ce virus est constitué d'un ARN positif d'environ 9500 bases qui code pour une polyprotéine de 3033 résidus [13], symbolisé sur la figure 1 annexé par le rectangle 1A. Cette polyprotéine est clivée après expression par des protéases endogènes et exogènes pour donner naissance à 10 protéines différentes. Deux d'entre elles, appelées El et E2, sont glycosylées et forment l'enveloppe du virus. Elles possèdent chacune des domaines membranaires appelés TM, en particulier TME1 pour la protéine E1 et TME2 pour la protéine E2. Les positions de clivage qui les génèrent

10

15

20

sont indiquées sur la figure 1 par des flèches avec, mentionné au-dessous, un nombre qui correspond à la position dans la polyprotéine du premier acide aminé de séquence résultant du clivage. Les protéines E1 et E2 sont symbolisées par un rectangle. La partie blanche de chaque rectangle correspond à l'ectodomaine (ed) et la partie grisée à la région transmembranaire (TM). La séquence primaire des TM est indiquée dans le bas de la une lettre, avec des figure en code nombres correspondant à la position des acides aminés dans la polyprotéine situés aux extrémités de ces domaines. Les étoiles indiquent les acides aminés hydrophobes. Ces domaines membranaires ou régions membranaires du virus ont des propriétés d'association particulières conditionnent la structuration de l'enveloppe du virus Α titre, ils constituent des cibles [12]. ce thérapeutiques potentielles. La compréhension mécanisme d'association du virus nécessite des études de la structure 3D de ces domaines notamment par les techniques précitées, ce qui implique de produire ces quantité abondante, peptides en et en de préférence par voie biosynthétique pour permettre un marquage isotopique 15N et/ou 13C.

Les différents essais d'expression de El 25 l'art antérieur, notamment dans E. coli [14][5] ou dans des cellules d'insecte sf9 infect ées par baculovirus [15], n'ont pas permis de surproduire cette protéine El, notamment du fait de la toxicité induite par son expression, compris dans les souches E . У coli 30 dites « résistantes » décrites BL21(DE3)pLysS dessus. Il n'y a pas eu d'essai de surexpression de la

20

25

protéine E2 dans la bactérie. Ces problèmes de toxicité sont essentiellement dus à la région C-terminale des deux protéines, riche en acides aminés hydrophobes qui forment des domaines transmembranaires assurant l'ancrage à la membrane du réticulum endoplasmique.

Il existe donc un réel besoin pour un système d'expression de protéines toxiques qui ne présente pas les inconvénients, limitations, défauts et désavantages des techniques de l'art antérieur.

10 En outre, il existe un réel besoin d'un vecteur d'expression comprenant un tel système d'expression de protéines toxiques permettant de mettre en œuvre un procédé de fabrication de protéines toxiques qui ne présente pas les inconvénients, limitations, défauts et désavantages des techniques de l'art antérieur.

#### EXPOSÉ DE L'INVENTION

Le but de la présente invention est précisément de fournir un système d'expression d'une protéine toxique, qui réponde, entre autres, aux besoins indiqués ci-dessus.

Ce but, et d'autres encore, sont atteints, conformément à l'invention, par un système d'expression caractérisé en ce qu'il comprend successivement, dans le sens 5'-3', une séquence nucléotidique codant pour le dipeptide Asp-Pro, appelée ci-dessous séquence dp, et une séquence nucléotidique (pt) codant pour une protéine toxique (Pt). Ce système sera identifié ci-dessous par : dp-pt.

30 Selon un mode particulièrement préféré de la présente invention, le système d'expression comprend en

15

20

25

30

outre, en amont de la séquence dp, une séquence nucléotidique (ps) codant pour une protéine soluble (Ps). Cette protéine soluble peut être par exemple la glutathion S-transférase (GST) ou la thiorédoxine (TrX) ou une autre protéine soluble équivalente. Ce système d'expression selon l'invention sera identifié cidessous par : ps-dp-pt.

Le système d'expression dp-pt de la présente invention, qui comprend une séquence codant pour Asp- ! Pro (DP en code à une lettre) placée en amont de la séquence nucléotidique de la protéine toxique, permet de manière tout à fait inattendue de supprimer l'effet toxique de la protéine pour la cellule hôte. En outre, les inventeurs ont noté que, de manière tout à fait surprenante, la suppression de toxicité de la protéine chez l'hôte est encore plus efficace avec le système d'expression ps-dp-pt, lorsque le peptide toxique est fusion C-terminale produit en avec une protéine soluble, par exemple la glutathion S-transférase ou la thiorédoxine, avec la séquence Asp-Pro insérée entre la protéine soluble et le peptide toxique.

Le système d'expression dp-pt ou ps-dp-pt de la présente invention permet la surproduction de protéines toxiques dans des cellules hôtes, en particulier de protéines hydrophobes, notamment de peptides qui correspondent à, ou qui comprennent, des domaines hydrophobes de protéines ancrées aux membranes. Il peut s'agir par exemple d'une protéine membranaire ou un domaine d'une protéine membranaire. Il peut s'agir par exemple d'une protéine d'un virus, par exemple d'un virus de l'hépatite C, d'un virus du sida, ou de tout

20

25

30

autre virus pathogène pour l'humain et de manière générale pour les mammifères.

Par exemple, le système dp-pt ou ps-dp-pt de l'invention permet de surproduire chez un hôte tel que E.coli les domaines transmembranaires des protéines El et E2 du virus de l'hépatite C, appelées TME1 et TME2, correspondant respectivement aux séquences :

TME1: 347-MIAGAHWGVLAGIAYFSMVGNWAKVLVVLLLFAGVDA-383

10 Séquence IDn°1

TME2: 717-MEYVVLLFLLLADARVCSCLWMMLLISQAEA-746 Séquence IDn°2

15 alors que cela n'était pas possible avec les techniques de l'art antérieur.

Les séquences nucléotidiques utilisables pour constituer le système dp-pt de l'invention codant pour les protéines TME1 (dp-pt(TME1)) ou TME2 (dp-pt(TME2)) peuvent être toutes les séquences possibles codant respectivement pour les protéines de fusion DP-TME1 et DP-TME2. Les séquences codant pour les protéines TME1 et TME2 peuvent être par exemple, avantageusement, respectivement, des séquences IDn°3 et IDn°4 de la liste de séquences annexée. Pour obtenir le système dp-pt la séquence dp codant pour le dipeptide Asp-Pro (DP) est ajoutée à ces séquences.

Les séquences nucléotidiques utilisables pour constituer le système ps-dp-pt de l'invention codant pour les protéines TME1  $(ps-dp-pt_{(TME1)})$  ou TME2  $(ps-dp-pt_{(TME2)})$  peuvent être toutes les séquences possibles

10

20

25

codant respectivement pour les protéines de fusion Ps-DP-TME1 et Ps-DP-TME2. Il peut s'agir par exemple, avantageusement, des séquences IDn°34, IDn°35 et IDn°36 de la liste de séquences annexée pour TME1, permettant d'obtenir une protéine chimère Ps-DP-TME1. Il peut s'agir par exemple, avantageusement, des séquences IDn°37, IDn°38 et IDn°39 de la liste de séquences annexée pour TME2, permettant d'obtenir une protéine chimère Ps-DP-TME2.

En effet, les séquences nucléotidiques précitées présentent des codons optimisés pour l'expression de TME1 et TME2 dans une bactérie, par exemple chez E. coli.

Il existe un grand nombre de séquences de l'ARN du

VHC produisant un phénotype infectieux : ces séquences sont
également utilisables dans la présente invention.

La séquence codant pour le dipeptide Asp-Pro peut être par exemple : gacccg, ou toute autre séquence codant pour ce dipeptide.

La séquence codant pour la GST peut être par exemple celle présente dans les plasmides pGEXKT dont la séquence correspond à la séquence IDn°29 de la liste de séquence annexée ou toute séquence équivalente, c'est à dire codant pour cette protéine soluble. La séquence codant pour la TrX peut être par exemple celle présente dans le plasmide d'expression pET32a+ dont la séquence correspond à la séquence IDn°30 de la liste de séquence annexée ou toute séquence équivalente, c'est à dire codant pour cette protéine soluble.

30 Pour la fabrication de la protéine toxique, le système d'expression dp-pt ou ps-dp-pt de l'invention est placé à l'intérieur d'une cellule hôte par exemple par clonage dans un plasmide approprié et ceci par les

15

20

25

30

techniques habituelles de transformation d'un hôte dans les techniques de recombinaison génétique.

Le plasmide dans lequel le système d'expression de la présente invention peut être cloné pour former ce vecteur sera choisi notamment en fonction de la cellule hôte. Il peut s'agir par exemple du plasmide pT7-7 (séquence IDn°33 de la liste de séquences annexée), d'un plasmide de la série pGEX (par exemple de séquence IDn°31 de la liste de séquences annexée), commercialisé par exemple par la société Pharmacia, ou un plasmide de la série pET32 (par exemple de séquence IDn°32 de la liste de séquences annexée), commercialisé par exemple par la société Novagen.

Les plasmides de la série pGEX et de la série pET32 seront avantageusement utilisés pour la mise en de la présente invention. En effet, ils comprennent déjà une séquence ps codant pour protéine soluble (Ps), respectivement pour glutathion S-transférase et pour la thiorédoxine. Aussi, avantageusement, le système dp-pt sera cloné dans ces plasmides en aval de cette séquence ps codant pour la protéine soluble.

La présente invention se rapporte donc également à un vecteur d'expression comprenant un système d'expression selon l'invention dp-pt ou ps-dp-pt; en particulier un vecteur comprenant un système d'expression dp-pt selon l'invention et la séquence oligonucléotidique du plasmide pT7-7, ou un vecteur comprenant un système d'expression ps-dp-pt selon l'invention et la séquence oligonucléotidique d'un plasmide pGEX ou d'un plasmide pET32.

10

15

20

25

30

Par exemple, les vecteurs d'expression de la présente invention, convenant à un hôte bactérien tel que *E. coli*, et permettant la surexpression de la protéine membranaire TME1 précitée, peuvent avoir avantageusement une séquence oligonucléotidique choisie parmi les séquences IDn°40 (avec pGEXKT), IDn°42 (avec pET32a+) et IDn°44 (avec pT7-7) de la liste de séquences annexée.

Par exemple, les vecteurs d'expression de la présente invention, convenant à un hôte bactérien tel que *E. coli*, et permettant la surexpression de la protéine membranaire TME2 précitée, peuvent avoir avantageusement une séquence oligonucléotidique choisie parmi les séquences IDn°41 (avec pGEXKT), IDn°43 (avec pET32a+) et IDn°45 (avec pT7-7) de la liste de séquences annexée.

En effet, les vecteurs d'expression précités présentent des codons optimisés pour l'expression des protéines chimères de la présente invention, incluant TME1 et TME2, dans une bactérie, par exemple chez E. coli.

La présente invention se rapporte également à une cellule procaryote transformée par un vecteur d'expression selon l'invention. Cette procaryote transformée par le vecteur d'expression de la présente invention doit de préférence permettre la surexpression de la protéine toxique pour laquelle code le vecteur. Ainsi, toute cellule hôte capable de d'exprimer le vecteur d'expression de la présente E.coli, invention utilisable, exemple est. par avantageusement, la souche E. coli BL21(DE3)pLysS.

La présente invention se rapporte également à un procédé de fabrication d'une protéine toxique par recombinaison génétique comprenant les étapes suivantes :

- 5 transformer une cellule hôte avec un vecteur d'expression selon l'invention,
  - cultiver la cellule hôte transformée dans des conditions de culture telles qu'elle fabrique une protéine de fusion comprenant le dipeptide Asp-Pro suivi de la séquence peptidique de la protéine toxique à partir dudit vecteur d'expression, et
    - isoler ladite protéine de fusion, et
    - cliver ladite protéine de fusion pour récupérer la protéine toxique.
- Les étapes de transformation, de culture, ainsi que de l'isolement de la protéine chimère fabriquée peuvent être réalisées par les techniques habituelles de recombinaison génétique, par exemple par des techniques telles que celles qui sont décrites dans le document [25].

L'étape consistant à isoler la protéine de fusion peut être réalisée par les techniques habituelles connues de l'homme du métier pour isoler une protéine d'un extrait cellulaire.

- La protéine de fusion fabriquée par le procédé de l'invention a une séquence « Protéine soluble-Asp-Pro-protéine toxique ». Dans la présente description, le dipeptide Asp-Pro est aussi appelé DP suivant le code à une lettre des acides aminés.
- 30 Par exemple, lorsque la protéine toxique est TME1, la protéine de fusion peut avoir la séquence

10

15

20

IDn°46 de la liste de séquences annexée, qui correspond à la protéine de fusion GST-DP-TME1; la séquence IDn°48 de la liste de séquences annexée, qui correspond à la protéine de fusion TrX-DP-TME1; ou la séquence IDn°50 de la liste de séquences annexée, qui correspond à la protéine de fusion M-DP-TME1 de la liste de séquences annexée.

Par exemple, lorsque la protéine toxique est TME2, la protéine de fusion peut avoir la séquence IDn°47 de la liste de séquences annexée, qui correspond à la protéine de fusion GST-DP-TME2; la séquence IDn°49 de la liste de séquences annexée, qui correspond à la protéine de fusion TrX-DP-TME2; ou la séquence IDn°51 de la liste de séquences annexée, qui correspond à la protéine de fusion M-DP-TME2 de la liste de séquences annexée.

L'étape de clivage de cette protéine de fusion peut être réalisée avantageusement au moyen d'acide formique qui clive la protéine de fusion au niveau du dipeptide Asp-Pro. Elle peut être réalisée par ailleurs par toute technique appropriée connue de l'homme du métier pour récupérer une protéine d'un échantillon à partir d'une protéine de fusion.

Les inventeurs sont les premiers à avoir trouvé un système réellement efficace pour produire et même de surproduire, notamment dans la bactérie Escherichia coli (E. coli) des peptides hydrophobes correspondant aux domaines membranaires des protéines El et E2 de l'enveloppe du virus de l'hépatite C, dont l'expression est létale pour le microorganisme.

10

15

20

25

30

14

d'applications de présente Le champ la invention concerne principalement la production peptides hydrophobes à grande échelle notamment pour la recherche fondamentale et industrielle. En outre, production de la protéine chimère constituée de la protéine soluble, du dipeptide Asp-Pro, et du peptide hydrophobe peut être utilisée dans un but fonctionnel, notamment pour renseigner sur le degré d'oligomérisation du domaine membranaire ou encore sur sa capacité d'hétéropolymérisation.

Les protéines de fusion, ou protéines chimères, sont produites via leur ADN codant présent par exemple dans des plasmides commerciaux et à la suite desquels est introduit en phase l'ADN codant pour la séquence Asp-Pro suivi de celui codant pour le peptide toxique. Cette application peut être commercialisée sous de plasmides d'expression bactériens incluront la séquence du site Asp-Pro, en aval de celle des protéines solubles déjà présentes. Le plasmide correspondant sera décrit par exemple comme un outil facilitant la production par voie biologique peptides ou protéines membranaires toxiques.

Ainsi, la présente invention est applicable à tout système de surexpression de protéines recombinantes, avec ou sans fusion à une protéine soluble comme par exemple la GST, ou la thiorédoxine, incluant une séquence Asp-Pro non naturelle insérée en amont d'une séquence codant pour un domaine toxique de la protéine, par exemple un domaine membranaire d'une protéine.

15

D'autres caractéristiques et avantages de la présente invention apparaîtront encore à l'homme du métier à la lecture des exemples suivants donnés à titre illustratif et non limitatif en référence à la liste de séquences et aux figures annexées.

# Brève description de la liste de séquences annexée

- Séquences IDn°1 et 2 : respectivement, séquences peptidiques de TME1 et de TME2.
- 10 Séquences IDn°3 et 4 : respectivement, séquences codant pour le peptide TME1 et le peptide TME2.
  - Séquences IDn°5 et 6 : respectivement, oligonucléotide (+) d'insertion dans pT7-7 (OL13(+)) et oligonucléotide (-) d'insertion dans pT7-7 (OL14(-)).
  - Séquences IDn°7 et 8 : respectivement, ADN sens codant de TME1 + site cla I en 3' et ADN sens anticodant de TME1 + site cla I en 5' (séquence complémentaire de la séquence IDn°7).
- 20 Séquences IDn°9 et 10 : respectivement, oligonucléotide sens codant (OL11(+)) et oligonucléotide sens anticodant (OL12(-)) pour la synthèse de TME1.
  - Séquence IDn°11 : oligonucléotide (+)
- 25 d'insertion dans pGEXKT sans site dp (OL15(+)).
  - Séquence IDn°12 : oligonucléotide (+) d'insertion dans pGEXKT avec site dp (OL17(+)).
  - Séquence IDn°13 : oligonucléotide (-) d'insertion dans pGEXKT (OL16(-)).

16

- Séquences IDn°15 et 16 : respectivement, oligonucléotides (+) (OL19(+)) et (-) (OL20(-)) d'insertion dans pT7-7 de l'ADN codant pour MDP-TME1.
- 5 Séquences IDn°17 et 18 : respectivement, oligonucléotide (+) d'insertion dans pT7-7 (OL23(+)) et oligonucléotide (-) d'insertion dans pT7-7 (OL24(-)).
- Séquences IDn°19 et 20 : respectivement, ADN sens codant pour TME2 + site de Nde I en 5' et Hind III en 3'; et ADN sens anticodant de TME2 + site Nde I en 3' et Hind III en 5' (séquence complémentaire de IDn°17).
- Séquences IDn°21 et 22 : respectivement,

  15 oligonucléotides sens codant (OL21(+)) et
  anticodant (OL22(-)) pour la synthèse de TME2.
  - Séquence IDn°23 : oligonucléotide (+) d'insertion dans pGEXKT sans site dp (OL25(+)).
- Séquences IDn°24 et 25 : respectivement, 20 oligonucléotides (+) (OL27(+)) et (-) (OL26(-)) d'insertion dans pGEXKT avec site dp.
  - Séquences IDn°26 et 27 : respectivement, oligonucléotides (+) (OL28(+)) et (-) (OL29(-)) d'insertion dans pT7-7 de l'ADN codant pour MDP-TME2.
  - Séquence IDn°28 : fin de la séquence de la protéine soluble GST suivie du site thrombine codée dans le plasmide pGEXKT.
- Séquence IDn°29 : ADN codant pour la protéine 30 GST dans le plasmide pGEXKT.

20

25

17

- Séquence IDn°30 : ADN codant pour la thiorédoxine (TrX) dans le plasmide pET32a+.
- Séquences IDn°31, 32 et 33 : respectivement, plasmides d'expression pGEXKT, pET32a+ et pT7-7.
- 5 Séquences IDn°34, 35 et 36 : respectivement, systèmes d'expression selon l'invention codant pour les protéines de fusion GST-DP-TME1, TrX-DP-TME1 et M-DP-TME1.
- Séquences IDn°37, 38 et 39 : respectivement,

  systèmes d'expression selon l'invention codant pour
  les protéines de fusion GST-DP-TME2, TrX-DP-TME2 et
  M-DP-TME2.
  - Séquences IDn°40 et 41 : respectivement, vecteurs d'expression pGEXKT-dp-pt<sub>TME1</sub> et pGEXKT-dp-pt<sub>TME2</sub> selon l'invention codant pour les protéines de fusion GST-DP-TME1 et GST-DP-TME2.
  - Séquences IDn°42 et 43 : respectivement, vecteurs d'expression pET32a-dp-pt<sub>TME1</sub> et pET32a-dp-pt<sub>TME2</sub> selon l'invention codant pour les protéines de fusion TrX-DP-TME1 et TrX-DP-TME2 (codent par le brin complémentaire).
    - Séquences IDn°44 et 45 : respectivement, vecteurs d'expression pT7-7-dp-pt<sub>TME1</sub> et pT7-7-dp-pt<sub>TME2</sub> selon l'invention codant pour les protéines de fusion M-DP-TME1 et M-DP-TME2.
    - Séquences IDn°46 et 47 : respectivement, protéines de fusion GST-DP-TME1 et GST-DP-TME2 selon l'invention obtenues à partir des plasmides pGEXKT-dp-pt<sub>TME1</sub> et pGEXKT-dp-pt<sub>TME2</sub>.
- 30 Séquences IDn°48 et 49 : respectivement, protéines de fusion TrX-DP-TME1 et TrX-DP-TME2 selon

15

30

l'invention obtenues à partir des plasmides pET32a- $dp-pt_{\text{TME1}}$  et pET32a- $dp-pt_{\text{TME2}}$ .

- Séquences IDn°50 et 51 : respectivement, protéines de fusion M-DP-TME1 et M-DP-TME2 selon l'invention obtenues à partir des plasmides pT7-7-dp-pt<sub>TME1</sub> et pT7-7-dp-pt<sub>TME2</sub>.
- Séquences IDn°52 et 53 : respectivement, protéines GST et TrX codées par le vecteur pGEXKT et pET32a+.

# 10 Brève description des figures

- Figure 1 : représentation schématique d'une partie de la polyproptéine du VHC et séquence peptidique des domaines membranaires C-terminaux des d'enveloppe E1 et E2. Les séquences peptidiques représentées correspondent au type infectieux #D00831 et #M67463 pour TME1 et TME2 respectivement obtenus de la banque publique de séquences Du Laboratoire Européen de Biologie Moléculaire (« European Molecular Biology Laboratory ») (EMBL).
- 20 Figure 2 : création de l'ADN codant pour le domaine membranaire C-terminal de la protéine d'enveloppe E1 du VHC et séquences additionnelles en 5' et 3' pour clonage dans différents plasmides. Les séquences représentées sur cette figure sont reportées dans la 25 liste de séquences annexée.
  - Figure 3 : création de l'ADN codant pour le domaine membranaire C-terminal de la protéine d'enveloppe E2 du VHC et séquences additionnelles en 5' et 3' pour clonage dans différents plasmides. Les séquences représentées sur cette figure sont reportées dans la liste de séquences annexée.

- Figure 4, panneaux A à F: toxicité des domaines membranaires exprimés dans la bactérie et suppression de cette toxicité par insertion d'un site dp. Les panneaux A, C et E sont des représentations graphiques de mesures de la densité optique (DO) à 600 nm en fonction du temps (t) en heures de la production de différentes protéines dans une bactérie utilisant ou non le système d'expression de la présente invention. Les panneaux B, D et F sont des représentations des gels de migration des protéines respectivement des panneaux A, C et E.
- Figures 5 A et B : surexpression des protéines thiorédoxine-Asp-Pro-Pt (Pt chimères domaines membranaires des protéines) dans la bactérie. La figure 5A est une représentation graphique de mesures de la 15 densité optique (DO) à 600 nm en fonction du temps en heures de la production de différentes protéines dans une bactérie utilisant ou non le système d'expression la présente invention. La figure 5B représentation d'un gel de migration des protéines de 20 la figure 5A.
  - Figure 6: expression et purification de la fusion chimère) GST-TME2, protéine de (ou comparaison avec la GST seule. Cette figure représente en haut les séquences peptidiques de GST et GST-TME2, et en bas les gels obtenus par électrophorèse montrant que contrairement à GST seule, GST-TME2 est insoluble. Cette dernière est produite sous forme d'inclusion qui ne peuvent pas se replier correctement.
- 30 Figures 7A et 7B : représentations graphiques de résultats expérimentaux comparatifs montrant l'effet du

dipeptide DP (séquence oligonucléotidique dp-pt conforme à la présente invention) et du dipeptide DP et de la protéine soluble (séquence oligonucléotidique ps-dp-pt conforme à la présente invention) sur la synthèse des protéines toxiques TME1 et de TME2 conformément à la présente invention.

#### **EXEMPLES**

10 Dans ces exemple, les oligonucléotides utilisés ont été commandés aux Laboratoires EUROBIO (http://www.eurobio.fr/); les plasmides ont préparés avec le kit QlAprep (marque de commerce) de QIAGEN (http://www.qiagen.com/); les séquences d'ADN 15 ont été séquencées avec le kit ABI PRISM (marque déposée) BigDye (marque de commerce) Terminator cycle d'Applied Biosystems (http://home.appliedbiosystems.com/); les souches d'E. coli BL21(DE3) et BL21(DE3)pLysS ont été obtenues chez 20 Stratagene (http://www.stratagene.com/); les souches C41 et C43 (BL21(DE3)) ont été fournies par le Dr. Bruno Miroux (CNRS-CEREMOD, Centre de recherches sur l'endocrinologie moléculaire et le développement ; les enzymes de restriction et modification d'ADN ont été 25 obtenues chez New England Biolabs (http://www.neb.com/neb/); les électrophorèses protéines ont été réalisées avec un miniprotean 3 (marque de commerce) Laboratoires Bio-Rad des (http://www.bio-rad.com); le plasmide pCR (marque 30 déposée) T7 topo TA a été obtenu chez Invitrogen (http://www.invitrogen.com/); le plasmide pET32a+ a été

15

25

21

obtenu chez Novagen (http://www.novagen.com); les plasmides pT7-7 et pGP1-2 et la souche K38 [22] ont été demandés au Prof. Tabor (Department of Biological Chemistry, Harvard Medical School); le plasmide pGEX-KT a été demandé au Prof. Dixon (Department of Biological Chemistry, University of Michigan Medical School); les autres produits ont été obtenus chez Sigma (http://sigma.aldrich.com).

Dans les exemples suivants, la production des peptides TME1 et TME2 a été dans un premier temps réalisée sans le système d'expression de la présente invention, puis en fusion avec une protéine soluble, et enfin en fusion avec la GST avec insertion du site Asp-Pro (« DP » en codage à une lettre) entre la protéine soluble et TME1 ou TME2.

L'abréviation « SEQ IDn° » est utilisée pour « séquence IDn° » et renvoie à la liste de séquences annexée.

# 20 Exemple 1 : Synthèse du système d'expression

# 1.1) CONSTRUCTION DES VECTEURS D'EXPRESSION pT7-7-pt<sub>TME1</sub> ET pT7-7-pt<sub>TME2</sub>

L'ADN codant pour les deux domaines a été synthétisé de novo à partir des oligonucléotides appropriés. Les codons ont été choisis suivant leur plus grande fréquence d'utilisation dans la bactérie, telle qu'elle a été quantifiée par Sharp et col. [17]. Les constructions sont décrites sur la figure 2 annexée pour TME1 et sur la figure 3 annexée pour TME2.

30 Chaque ADN synthétique a été généré en utilisant un jeu de deux oligonucléotides longs et

25

22

chevauchants, OL11(SEQ IDn°9) et OL12 (SEQ n°10) pour TME1 et OL21(SEQ IDn°19) et OL22 (SEQ IDn°20) pour TME2, qui ont été amplifiés après hybridation avec deux oligonucléotides externes choisis en fonction du clonage dans un plasmide donné. Ainsi les clonages dans pT7-7 ont été réalisés en utilisant le jeu d'oligonucléotides externes OL13 (SEQ IDn°5) et OL14 (SEQ IDn°6) pour TME1 et OL23 (SEQ IDn°15) et OL24 (SEQ IDn°16) pour TME2.

10 Chaque ADN synthétique a été généré en utilisant un jeu de quatre oligonucléotides : deux longs et chevauchants et deux courts et externes. Les ADN ont été amplifiés par la méthode de réaction de polymérisation en chaîne (« polymérase chain reaction ») dite de "PCR" [18] puis clonés dans un plasmide bactérien pCR (marque de commerce) T7 topo TA. Les ADN synthétisés ont été séquencés puis sous-clonés dans le vecteur d'expression bactérien pT7-7 [19] en utilisant les sites de restriction Nde I en 5' et Cla I ou Hind III en 3'.

# Sur la figure 2 :

- A : Séquence peptidique TME1 du sous-type #D00831. La numérotation correspond à la position de la séquence dans la polyprotéine comme décrit sur la figure 1.
- B : Séquence d'ADN codant pour le domaine membranaire avec des codons optimisés pour l'expression dans la bactérie.
- C et D: Stratégie d'amplification sans matrice de 30 l'ADN. Le sens codant et le sens anticodant des oligonucléotides sont indiqués respectivement par les

signes (+) et (-). Les oligonucléotides longs sont chevauchants d'une vingtaine de bases pour créer l'amorce puis la matrice. Les oligonucléotides courts permettent d'amplifier la matrice par PCR en intégrant 5 sites de restriction désirés et plasmides utilisés. L'insertion dans pT7-7 a réalisée avec le couple d'oligonucléotides OL13 (SEQ IDn°5) et OL14 (SEQ IDn°6), via un sous-clonage dans pCRT7 topo, en intégrant les sites Nde I et Hind III. 10 L'insertion dans pGEXKT a été effectuée suivant la même méthode avec le couple d'oligonucléotides OL15 (SEQ IDn°11) et OL16 (SEQ IDn°13) en intégrant les sites BamH I et EcoR I. L'insertion du site dp (gacccg) et le clonage dans pGEXKT ont été effectués avec le couple 15 d'oligonucléotides OL17 (SEQ IDn°12) et OL16 IDn°13). La construction dans pGEXKT a été transférée dans pET32a qui code pour la thiorédoxine avec le couple d'oligonucléotides OL18 (SEQ IDn°14) et OL16 IDn°13). L'oligonucléotide OL18 (SEQ IDn°14) s'hybride dans la région terminale de l'ADN codant pour 20 la GST dans pGEXKT. La séquence amplifiée intègre la fin de la GST (SDLSGGGGG) suivie du site thrombine (LVPRGS) (SEQ IDn°28), du site DP et du passage membranaire. Après clonage, l'ADN inséré dans pET32a 25 chimère thiorédoxinepermet d'exprimer la. SDLSGGGGGLVPRGS-DP-TME1 (SEQ IDn°48).

# Sur la figure 3 :

La légende est identique à la figure 2, mais La 30 séquence peptidique est celle du sous-type #M67463.

L'insertion dans pT7-7 a été réalisée avec le couple

d'oligonucléotides OL23 et OL24 (respectivement SEQ IDn°17 et SEQ IDn°18) via un sous-clonage dans pCRT7 topo en intégrant les sites Nde I et Hind III.

L'insertion dans pGEXKT a été effectuée suivant la même méthode avec le couple d'oligonucléotides OL25 et OL26 (respectivement SEQ IDn°23 et SEQ IDn°25) intégrant les sites BamH I et EcoR I. L'insertion du site dp (gacccg) et le clonage dans pGEXKT ont été effectués avec le couple d'oligonucléotides OL27 et OL26 (respectivement SEQ IDn°24 et SEQ IDn°25). L'insertion dans pET32a a été effectuée comme décrit la figure 2 utilisant en d'oligonucléotides OL18 et OL26 (respectivement SEQ IDn°14 et SEQ IDn°25).

15

20

25

30

10

# 1.2) CONSTRUCTION DES VECTEURS D'EXPRESSION pGEXKT-pt<sub>TME1</sub>, pGEXKT-pt<sub>TME2</sub>, pGEXKT-dp-pt<sub>TME1</sub> ET pGEXKT-dp-pt<sub>TME2</sub>

Les vecteurs d'expression pGEXKT-pt<sub>TME1</sub> et pGEXKT-pt<sub>TME2</sub> ont été construits par PCR comme décrits sur les figures 2 et 3 annexées. L'ADN-matrice utilisé pour amplifier les ADN codant pour TME1 ou TME2 est celui cloné dans les plasmides pT7-7. Le clonage dans le plasmide pGEXKT [20, 21] de TME1 a été réalisé en utilisant les jeux d'oligonucléotides OL15 (SEQ IDn°11) et OL16 (SEQ IDn°13) permettant l'insertion des sites de restriction BamH I en 5' et EcoR I en 3'. Le clonage de TME2 dans le même vecteur a été réalisé en utilisant les jeux d'oligonucléotides OL25 (SEQ IDn°21) et OL26 (SEQ IDn°23).

20

25

30

Comme indiqué dans la figure 2, l'insertion du site dp en position N-terminale de TME1 a été réalisée en remplaçant l'oligonucléotide 5' OL15 (SEQ IDn°11) par l'oligonucléotide OL17 (SEQ IDn°12). L'insertion du site dp en position N-terminale de TME2 a été réalisée en remplaçant l'oligonucléotide 5' OL25 (SEQ IDn°21) par l'oligonucléotide OL27 (SEQ IDn°22), comme montré dans la figure 3.

# 10 1.3) CONSTRUCTION DES VECTEURS D'EXPRESSION pET32a-dp-TME1 ET pET32a-dp-TME2

Les vecteurs d'expression pET32a-dp-TME1 pET32a-dp-TME2 ont été construits par PCR comme décrits sur les figures 2 et 3 annexées, en utilisant le jeu d'oligonucléotides indiqué. L'oligonucléotide amont intègre un site EcoR V et s'hybride avec la région terminale de gène codant pour la GST. Il permet d'intégrer la queue de 5 glycines et le site de clivage la thrombine présents dans le plasmide. L'oligonucléotide en aval est le même que celui utilisé pour le clonage dans pGEXKT.

L'insertion dans le plasmide pET32a se fait par les sites MsC I/EcoR V en 5' et EcoR I en 3'. Il permet d'insérer en phase à la fin de la séquence de la thiorédoxine la queue de 5 glycines, le site de clivage par la thrombine, le site DP et le passage membranaire. Le plasmide d'origine pET32a, qui sert de témoin, code pour la thiorédoxine suivie d'une séquence intégrant divers éléments qui n'ont pas été supprimés et qui contribuent pour une bonne part à la masse de la protéine chimère fabriquée.

L'ADN-matrice utilisé pour amplifier les ADN codant pour TME1 ou TME2 est celui cloné dans les plasmides pGEXKT-dp-pt<sub>TME1</sub> ou pGEXKT-dp-pt<sub>TME2</sub>. Pour TME1, le clonage dans pET32a+ a été réalisé en utilisant les jeux d'oligonucléotides OL18 (SEQ IDn°14) et OL16 (SEQ IDn°13). Le clonage de TME2 dans le même vecteur a été réalisé en utilisant les jeux d'oligonucléotides OL18 (SEQ IDn°14) et OL26 (SEQ IDn°23), comme indiqué dans la figure 3.

10.

15

20

25

30

5

# Exemple 2 : Expression de séquences codant pour les protéines TME1 et TME2 seules

L'expression des séquences codant pour les domaines seuls TME1 et TME2 a été testée par induction thermique ou chimique et en utilisant différentes souches bactériennes comme décrit ci-après.

## 2.1) SYSTEME D'INDUCTION THERMIQUE

Le système développé par Tabor [22] permet d'exprimer une protéine par induction thermique en utilisant deux vecteurs dans une même bactérie, pT7-7 et pGP1-2.

Le plasmide pT7-7 contient l'ADN à exprimer, placé sous contrôle d'un promoteur  $\phi 10$  reconnu par l'ARN polymérase du phage T7. Le plasmide pGP1-2 contient le gène codant pour la polymérase du phage T7, placé sous le contrôle d'un promoteur  $\lambda p_L$ . Ce promoteur est réprimé par un répresseur thermosensible, cI857, présent lui aussi dans pGP1-2. A 30°C, cI857 est normalement exprimé et réprime le promoteur  $\lambda p_L$ , ce qui

bloque l'expression de la polymérase et donc aussi celle de la protéine d'intérêt.

L'induction est déclenchée par un passage de la culture de 37 à 42°C pendant 15-30 min, puis l'expression se poursuit à 37°C. Ce système est donc particulièrement bien adapté lorsqu'il est nécessaire de contrôler strictement l'expression d'une protéine donnée, notamment si celle-ci est toxique pour la bactérie.

10

15

20

25

30

5

# 2.2) SYSTEME D'INDUCTION CHIMIQUE

plasmide pT7-7 contenant l'ADN à Le même exprimer est cette fois introduit dans des bactéries E. coli de type BL21(DE3) (B F dcm omtP hsdS( $r_B^-m_B^-$ ) gal  $\lambda$ (DE3) et BL21(DE3)pLysS (B F dcm ompT hsdS( $r_B^-m_B^-$ ) gal  $\lambda$ (DE3) [pLysS Cam<sup>r</sup>]). Ces bactéries ont été modifiées pour contenir dans le génome une copie du gène codant pour l'ARN polymérase du phage T7, placé sous contrôle d'un promoteur lacUV5 inductible par l'isopropyle-1thio- $\beta$ -D-galactoside (IPTG). Dans ce cas, les bactéries sont cultivées à leur température optimum de 37°C ou moins si nécessaire. L'expression est induite par addition d'IPTG dans la culture. La BL21(DE3)pLysS est particulièrement bien adaptée aux protéines dont l'expression basale est toxique pour la bactérie hôte. En effet, la présence du plasmide pLysS permet l'expression continue et à un faible niveau de lysozyme du phage T7. Celui-ci inhibe la polymérase du dont la faible expression en absence т7, d'induction pourrait permettre l'expression basale de protéine toxique.

Les inventeurs ont aussi testé l'expression des domaines membranaires seuls dans des souches appelées C41 et C43 [10] qui ont été sélectionnées pour résister à l'expression de protéines membranaires toxiques. Ces souches sont dérivées de la souche BL21(DE3) et sont utilisées de la même façon que cette dernière.

# 2.3) TESTS D'EXPRESSION

Suivant le système testé, les plasmides correspondants ont été introduits par transformation dans les différentes souches  $d'E.\ coli$ : K38 (HfrC  $\lambda$ ) pour le système d'induction thermique de Tabor ou les différentes souches BL21 pour l'induction chimique. Le tableau 1 suivant résume les tests effectués.

15

10

Tableau 1

Induction	Souche	Plasmide	
Thermique	K38	pT7-7 + pGP1-2	
Chimique	BL21(DE3)	pT7-7	
Chimique	BL21(DE3)pLysS	pT7-7	
Chimique	C41 (BL21(DE3))	pT7-7	
Chimique	C43 (BL21(DE3))	pT7-7	
	1	1	

Dans chaque cas, une dizaine de transformants 20 ont été mis en culture pour tester l'expression. Brièvement, les bactéries ont été mises en culture dans 5 ml de LB (10 g tryptone, 5 g extrait de levure, 5 g NaCl, qsp 1 litre H<sub>2</sub>O) complémenté avec 50 µg/ml d'ampicilline (nécessaire pour maintenir pT7-7 dans la bactérie) et 60 µg/ml de kanamycine (nécessaire pour

10

15

20

maintenir pGP1-2 dans la bactérie) puis cultivées jusqu'à saturation soit à 30°C pour K38, soit à 37°C pour BL21(DE3). Les cultures sont ensuite diluées au 1/10 dans le même milieu de culture et cultivées jusqu'à une densité optique (DO) de 1 mesurée à 600 nm sur un spectrophotomètre Philips PU8740 (marque de commerce).

L'expression est ensuite induite soit thermiquement (K38) à 42°C pendant 15 min, soit chimiquement (BL21(DE3)) par addition de 1 mM d'IPTG. Elle est poursuivie pendant 3-5 heures à 37°C. La DO<sub>600nm</sub> des cultures est mesurée à différents temps.

A la fin de l'expression, un volume de culture contenant l'équivalent de 0,1 DO de bactéries est prélevé. Les bactéries sont récoltées par centrifugation et suspendues dans 50  $\mu$ l de solution de lyse (SL: Tris-Cl 50 mM, pH 8,0, EDTA 2,5 mM, SDS 2%, urée 4 M,  $\beta$ -mercaptoéthanol 0,7 M). Après quelques minutes à température ambiante, 10  $\mu$ l sont déposés sur gel de polyacrylamide 16,5% pour électrophorèse de type "Tricine" [23] qui permet de bien séparer les protéines de petite masse moléculaire.

## Sur la figure 4 :

Panneaux A, C et E: Les bactéries ont été transformées avec les plasmides pT7-7, pT7-7-TME1, pT7-7-TME2 (panneau A), pGEXKT, pGEXKT-TME1, pGEXKT-TME2 (panneau C) et pGEXKT-dp-TME1, pGEXKT-dp-TME2 (panneau E) puis mises en culture et induites comme décrit ci-dessus. La croissance bactérienne a été suivie en mesurant l'augmentation de turbidité de chaque culture par

10

15

20

25

30

mesure de la densité optique à 600 nm en fonction du temps en heures.

Panneaux B, D, F: Les bactéries ont été prélevées au temps indiqué dans le texte et traitées comme décrit ci-dessus. Elles ont ensuite été déposées sur gel d'électrophorèse, soit 16,5% d'acrylamide de type "Tricine" (panneau B), soit 14% d'acrylamide de type Laemmli SDS-PAGE (panneaux D et F). L'électrophorèse montrée dans le panneau F a migré plus longtemps que celle montrée dans le panneau D, ceci pour améliorer la séparation des bandes dans la zone des 30000 Da. Après migration, les gels ont été colorés 10 minutes par du bleu de Coomassie dans une solution de 40% méthanol, 10% d'acide acétique, 0,1% bleu de Coomassie R250, puis décolorés dans une solution 10% méthanol, 10% acide acétique, 1% glycérol.

Quelque soit le système testé, la première observation est que la fréquence de transformation des bactéries a été faible. Pour les bactéries qui ont pu être sélectionnées, le résultat des tests d'expression a été systématiquement négatif. Un exemple est donné sur la figure 4, panneaux A et B, avec les séries BL21(DE3)pLysS {[pT7-7], [pT7-7-TME1] ou [pT7-7-TME2]}. l'illustre la comparaison des courbes croissance du panneaux A de la figure 4, les inventeurs constaté avec les clones transformés pT7-7-TME1 ou pT7-7-TME2 et résistants sur milieu solide, que l'induction stoppe quasi-immédiatement la croissance bactérienne, contrairement aux contenant le plasmide seul. De même, comme c'est

visible sur la figure 4(B), aucune bande de protéines migrant dans la région correspondant à la masse moléculaire des produits d'expression (~ 3-4000 Da) ou d'oligomères de ceux-ci ({1, 2, 3, ...}) x masse moléculaire) n'est en effet observable.

la L'explication plus probable à cette l'expression situation est que des domaines membranaires est très toxique pour la bactérie. difficulté d'obtenir des transformants laisse supposer qu'une expression basale, même très faible, suffit à les tuer. Ceci montre aussi que le système pLysS n'est pas parfait pour prévenir cette expression basale. les bactéries qui résistent à l'étape de transformation, l'induction de l'expression domaines hydrophobes devient immédiatement létale. Les systèmes utilisés permettent effectivement de protéger la bactérie hôte d'une expression basale, mais dès que celle-ci est induite, la toxicité est immédiate et les bactéries sont tuées.

20

25

10

15

# Exemple 3 : Expression de séquences codant pour les protéines de fusion GST-TME1 et GST-TME2

Les vecteurs d'expression ont été construits comme décrit dans l'exemple 1, puis introduits dans les bactéries BL21(DE3)pLysS. Les bactéries BL21(DE3)pLysS ont été utilisées par soucis de comparaison avec les expériences précédentes car l'expression de la GST ou de ses chimères ne nécessite pas le système DE3-pLysS.

L'induction de l'expression a été effectuée 30 avec de l'IPTG comme pour celle des domaines seuls. Les

caractéristiques des protéines produites sont résumées dans le tableau 2 suivant.

Tableau 2

Plasmide	Chimère, abréviation	Construction	Taille, aa	Masse, Da
pGEXKT	GST, G	1M-D <sub>239</sub>	239	27469
pGEXKT-T1	GST-TME1, GT1	1M-S <sub>233</sub> -347M-A <sub>383</sub>	269	30506
pGEXKT-T2	GST-TME2, GT2	<sub>1</sub> M-S <sub>233-717</sub> E-A <sub>746</sub>	263	30191

5

10

15

20

Les acides aminés (aa) sont indiqués avec le code à une lettre. La numérotation des séquences est effectuée par rapport aux protéines d'origine, GST et polyprotéine virale. Celle qui fait référence aux domaines membranaires est indiquée en italiques.

Les panneaux C et D de la figure 4 annexée les résultats obtenus. Les montrent courbes de des bactéries croissance transformées avec les différents plasmides montrent que l'expression chimères GT1 et GT2 est toxique. Comme on peut le voir sur le gel d'électrophorèse de type Laemmli SDS-PAGE 14% [24], l'expression de TME1 fusionné à la GST s'accompagne de l'absence de bande migrant à la taille attendue de 30 kDa. Ceci laisse supposer qu'un très faible niveau d'expression de la chimère est suffisant . pour tuer les bactéries. Par contre, la chimère GST-TME2 est cette fois visible sur d'électrophorèse, dans la zone de masse moléculaire

15

20

attendue de 30 kDa. Le niveau d'expression reste toutefois limité.

La protéine produite n'est pas soluble malgré la présence de la GST dans la fusion. En effet, comme le montre la figure 6 annexée, les essais de solubilisation, repliement et purification de la chimère GST-TME2 ont été un échec.

Pour obtenir les résultats représentés sur cette figure 6, les protéines GST et GST-TME2 ont été exprimées comme décrit sur la figure 4 à partir de 150 ml de milieu de culture. Les bactéries ont ensuite été récoltées par centrifugation et suspendues KPO<sub>4</sub> 20 mM pH 7.7, NaCl 0.1M, EDTA 1 mM, NaN<sub>3</sub> 1 mM) pour avoir 100 DO/ml. Deux ml de chaque culture sont prélevés pour sonication par pulses de 30 sec à 15% d'amplitude. Après sonication, Un prélèvement est effectué pour électrophorèse. Il correspond au puits « To » sur la figure 6 (correspondant au « total »).

Une première centrifugation à basse vitesse (5000xg, 15 minutes) permet de séparer les bactéries non cassées et les corps d'inclusion des protéines solubles ou membranaires. Ces dernières se retrouvent dans le surnageant et un prélèvement est effectué. Il correspond au puits « Surn » sur la figure 6.

La fraction contenant la GST seule est ensuite traitée avec une résine d'affinité qui permet de fixer puis d'éluer spécifiquement cette protéine (puits « Af » du gel GST sur la figure 6).

La fraction contenant la protéine non soluble 30 GST-TME2 est traitée soit avec un détergent doux comme le triton X100 (TX100), en présence ou en absence de

10

15

20

25

34

NaCl, soit avec un détergent plus solubilisant mais plus déstructurant comme le sarkosyl, avant d'être à nouveau diluée dans du TX100 et passée sur résine d'affinité.

Les résultats de la figure 6 montrent que la GST est présente dans la fraction soluble, à la différence de la fusion GST-TME2, ce qui indique que cette dernière est insoluble. Le surnageant contenant la GST est passé sur une résine d'agarose-GSH capable de fixer la GST. Celle-ci est ensuite éluée avec un excès de GSH (puits indiqué « Af » du gel GST sur la figure 6).

Le culot contenant la fusion GST-TME2 n'est pas solubilisé en présence d'un détergent doux comme le TX100 (avec ou sans NaCl ajouté, puits « TX100 +/-NaCl » du gel GST-TME2) mais il peut être solubilisé avec un détergent plus agressif comme le sarkosyl. Cependant, après dilution de la protéine ainsi solubilisée dans du TX100, un détergent doux qui doit favoriser son repliement, la protéine n'est pas retenue sur la résine d'affinité, contrairement à la GST, ce qui suggère que la protéine de fusion ne peut être repliée.

Ces tests indiquent de façon claire que la protéine GST-TME2 est produite sous forme de corps d'inclusion que l'on ne peut pas replier correctement.

# Exemple 4 : Expression de vecteurs d'expression codant pour les protéines de fusion incluant un site Asp-Pro et un site GST

30 La construction des vecteurs a été réalisée comme décrit plus haut et pour les deux vecteurs codant

pour les protéines chimères GST-TME1 et GST-TME2, pour donner naissance aux vecteurs codant pour les protéines chimères GST-Asp-Pro-TME1 et GST-Asp-Pro-TME2. Elles sont résumées dans le tableau 3 ci-dessous.

5

10

15

20

Tableau 3

Plasmide	Chimère ; Abréviation fig. 4	Construction	Taille, aa	Masse, Da
pGEXKT- dp-TI	GST-DP-TME1 ; G <sub>DP</sub> T1	<sub>1</sub> M-S <sub>233</sub> dp- <sub>347</sub> M- A <sub>383</sub>	271	30718
pGEXKT- dp-T2	GST-DP-TME2; G <sub>DP</sub> T2	<sub>1</sub> M-S <sub>233</sub> -dp- <sub>7/7</sub> E- A <sub>746</sub>	265	30403

Les acides aminés (aa) sont indiqués avec le code à une lettre. La numérotation des séquences est effectuée par rapport aux protéines d'origine, GST et polyprotéine virale. Celle qui fait référence aux domaines membranaires est indiquée en italiques.

Les vecteurs ont été testés comme décrit dans le paragraphe précédent. Les résultats obtenus sont illustrés sur les panneaux E et F de la figure 4 annexée.

Les courbes de croissance des bactéries transformées avec les différents plasmides montrent que l'expression des chimères  $G_{dp}T1$  et  $G_{dp}T2$  est nettément moins toxique que dans les cas précédents. Le panneau F montre que cette fois, TME1 est produite du fait de la présence du site de clivage DP. Son niveau d'expression, tel qu'on peut le voir dans le panneau F, est relativement moyen mais significatif. GST-DP-TME2

est nettement surproduit. Les deux protéines migrent dans leur zone de masse moléculaire attendue.

L'effet de l'addition du dipeptide DP est aussi significatif qu'inattendu: il amplifie l'expression des domaines et supprime leur toxicité. Cet effet d'atténuation de la toxicité n'est pas connu pour le dipeptide DP, dont la seule propriété rapportée à ce jour est sa capacité à être coupé par l'acide formique. L'effet étant observé sur deux peptides différents et tous deux initialement toxiques pour la bactérie, on peut donc raisonnablement penser que cette propriété peut s'étendre à d'autres peptides hydrophobes et toxiques.

Les inventeurs ont vérifié que le site peut 15 être effectivement clivé par l'acide formique : la coupure est lente et demande environ 7 jours à température ambiante.

Les essais d'expression à basse température  $(20\,^{\circ}\text{C})$  pendant toute une nuit de ces chimères ont permis de mettre en évidence qu'elles sont produites sous forme native. En effet, il est possible de détecter une activité GST transférase dans la fraction membranaire des bactéries. De plus, cette activité est mesurée en solution lorsque les membranes sont solubilisées en présence d'un détergent non ionique comme le  $\beta$ -D dodécylmaltoside, après centrifugation.

20

10

15

# Exemple 5 : Expression de vecteurs d'expression codant pour les protéines de fusion incluant un site Asp-Pro et un site codant pour la thiorédoxine (TrX)

Les vecteurs d'expression pET32a-TrX, pET32a-TrX-dp-TME1 et pET32a-TrX-dp-TME2 ont été construits comme décrit ci-dessus et ont été ensuite introduits dans les bactéries BL21(DE3)pLysS. Les bactéries BL21(DE3)pLysS ont été utilisées par soucis de comparaison avec les expériences précédentes car l'expression de la GST ou de ses chimères ne nécessite pas le système DE3-pLysS. Les clones positifs ont été mis en culture et induits comme décrit ci-dessus.

L'induction de l'expression a été effectuée avec de l'IPTG comme pour celle des domaines seuls. Les caractéristiques des protéines produites sont résumées dans le tableau 4 suivant.

Tableau 4\*

Plasmide	Chimère ; Abréviation fig. 4	Construction	Taille, aa	Masse, Da
pET32a	Thiorédoxine ; TrX	<sub>1</sub> M-C <sub>189</sub>	189	20397
pET32a-	TrX-DP-TME1;	<sub>1</sub> M-S <sub>115</sub> -PK-	171	17796
Gend-dp-T1	T <sub>DP</sub> T1	Gend-dp-T <sub>1</sub>		17790
рЕТЗ2а-	TrX-DP-TME2;	1M-S115-PK-	165	17481
Gend-dp-T2	T <sub>DP</sub> T2	Gend- <i>dp</i> -T₂	105	17461

\* : T1 = TME1 et T2 = TME2

20

Les acides aminés (aa) sont indiqués avec le code à une lettre. La numérotation des séquences est effectuée par rapport aux protéines d'origine, GST et polyprotéine virale. Celle qui fait référence aux

domaines membranaires indiquée en italiques. est « Gend » fait référence à la séquence C-terminale de la provenant des constructions avec le plasmide la séquence peptidique Elle correspond à pGEXKT. primaire SDLSGGGGGLVPRGS. Les chimères thiorédoxine-SDLSGGGGGLVPRGS-DP-(TME1 ou TME2) sont plus courtes que la protéine codée dans le vecteur d'origine l'insertion est effectuée immédiatement après thiorédoxine.

10.

5

### Sur la figure 5 :

A : la croissance bactérienne a été suivie en mesurant l'augmentation de turbidité de chaque culture par densité optique à 600 nm en fonction du temps.

15 **B**: les bactéries ont été prélevées comme indiqué pour la figure 4. Elles ont ensuite été déposées sur gel d'électrophorèse 14% d'acrylamide de type Laemmli SDS-PAGE et traitées comme indiqué pour la figure 4.

Comme attendu, et comme le montrent les courbes de croissance représentées sur la figure 5A annexée des bactéries transformées avec les différents plasmides, l'expression des chimères TrX-DP-TME1 et TrX-DP-TME2 suivant la présente invention n'est pas toxique. Le gel d'électrophorèse représenté sur la figure 5B annexé de type Laemmli SDS-PAGE 14% [24] montre que chaque chimère est surproduite.

La présente invention permet donc de fabriquer 30 par recombinaison génétique des peptides hydrophobes correspondant aux domaines membranaires des protéines

El et E2 de l'enveloppe du virus de l'hépatite C, dont l'expression a été reconnue comme létale dans les techniques de l'art antérieur. En outre, l'effet étant observé sur deux peptides vraiment différents et tous deux initialement toxiques pour la bactérie, indique que la présente invention concerne d'autres peptides hydrophobes et toxiques.

# Exemple 6 : Effet du dipeptide DP sur la toxicité des domaines transmembranaires TME1 et TME2 exprimés sans protéine de fusion dans la bactérie

Cet exemple permet d'évaluer l'effet antitoxique du dipeptide DP inséré en absence de protéine de fusion GST ou TrX conformément à la revendication 1 annexée.

15

20

25

A) Matériels: Les plasmides pT7-7-pt<sub>TME1</sub> et pT7-7-pt<sub>TME2</sub> sont ceux qui sont décrits dans l'exemple 1. Les plasmides pT7-7-dp-pt<sub>TME1</sub> et pT7-7-dp-pt<sub>TME2</sub> ont été construits et clonés dans pT7-7 (SEQ IDn°33) comme décrits dans l'exemple 1 mais en utilisant les sites Ι (3′) du plasmide. (5') et EcoR Ι oligonucléotides amont (5') intègrent la séquence dp (gacccg) après la lère méthionine (atg). Les matrices utilisées pour générer chaque ADN ont été les plasmides  $pT7-7-pt_{TME1}$  et  $pT7-7-pt_{TME2}$ . Les séquences ont été vérifiées après clonage.

Les oligonucléotides utilisés sont les suivants:

i) Clonage de la séquence codant pour (M)DP-30 TME1 dans pT7-7:

OL19 (+): 5'- CG CATATGGACCCGATCGCTGGTGCT -3'
(Nde I souligné) = (SEQ IDn°15 de la liste de séquences annexée);

OL20 (-): 5'-GAATTCCTAAGCGTCAACACCAGC-3' (EcoR

5 I souligné) = (SEQ IDn°16 de la liste de séquences annexée).

ii) Clonage de la séquence codant pour (M) DP TME2 dans pT7-7:

OL28 (+) : 5'- CGCATATGGACCCGGAATACGTTGTTC-3'

(Nde I souligné) = (SEQ IDn°26 de la liste de séquences annexée);

OL29 (-): 5'-CAGAATTCCTAAGCTTCAGCCTGAGAG-3'
(EcoR I souligné) = (SEQ IDn°27 de la liste de séquences annexée).

15

Les vecteurs d'expressions pT7-7-dp- $pt_{\rm TME1}$  et pT7-7-dp- $pt_{\rm TME2}$  obtenus sont présentés sur la liste de séquences annexée (SEQ IDn°44 et SEQ IDn°45).

souche bactérienne BL21(DE3)pLysS a été transformée soit avec le plasmide seul soit avec les différentes versions de pT7-7 intégrant les 4 constructions exprimant TME1, M-DP-TME1 (figure 7A) ou TME2, M-DP-TME2 (figure 7B). M représente la méthionine, elle est présente en position N-terminale des peptides lorsque les protéines toxiques sont fabriquées suivant la présente invention avec le plasmide pT7-7.

La croissance des différents clones a été 30 comparée après induction par l'IPTG, suivant le protocole identique à l'induction chimique décrite dans

41

l'exemple 2, et moyennée sur les valeurs de DO de 4 clones différents pour chaque construction.

### C) Résultats:

Les figures 7A et 7B montrent que les bactéries ayant un plasmide exprimant des protéines TME1 et TME2 croissent moins vite après induction que la souche témoin qui est transformée avec le vecteur pT7-7 seul.

10. Ces résultats montrent que les souches transformées avec les plasmides exprimant les versions M-DP-TME1 (SEQ IDn°50) et M-DP-TME2 (SEQ IDn°51) selon l'invention croissent significativement mieux que celles qui expriment les TM sans DP. Ceci est vrai pour 15 TME1 et encore plus nettement pour TME2.

La conclusion est que l'insertion N-terminale de DP conformément à la présente invention contribue de manière surprenante à une diminution significative de la toxicité de l'expression des domaines membranaires, notamment en absence d'une protéine de fusion soluble comme la GST ou la thiorédoxine.

42

#### Liste des références

- [1] Christendat D., Yee A., Dharamsi A., Kluger Y., Gerstein M., Arrowsmith C.H., and Edwards A.M., (2000), Prog. Biophys. Mol. Biol. 73, 339-345;
  - [2] Hammarstrom M., Hellgren N., Van Den Berg S., Berglund H., and Hard T., (2002), Protein Sci. 11, 313-321;
  - [3] Falson P. (1992), Biotechniques 13, 20-22;
- 10 [4] Falson P., Penin F., Divita G., Lavergne J.P., Di
  Pietro A., Goody R.S., and Gautheron D.C. (1993),
  Biochemistry 32, 10387-10397;
- [5] Ciccaglione A.R., Marcantonio C., Costantino A., Equestre M., Geraci A. and Rapicetta M. (2000) Virus 15 Genes 21, 223-226;
  - [6] Sisk W.P., Bradley J.D., Kingsley D., and Patterson
    T.A. (1992) Gene 112, 157-162;
  - [7] Paulsen I.T., Sliwinski M.K., Nelissen B., Goffeau A., and Saier M.H. Jr. (1998) FEBS Lett 430, 116-125;
- 20 [8] Decottignies A. and Goffeau A. (1997) Nat Genet 15, 137-145;
  - [9] Arechaga I., Miroux B., Karrasch S., Huijbregts R., de Kruijff B., Runswick M.J. and Walker J.E. (2000) FEBS Lett 482, 215-219;

- [10] Miroux B. and Walker J.E. (1996) J. Mol. Biol. 260, 289-298;
- [11] Mayo M.A., and Pringle C.R. (1998) J. Gen Virol. 79 (Pt4), 649-657;
- [12] Op De Beeck A., Montserret R., Duvet S., Cocquerel L., Cacan R., Barberot B., Le Maire M., Penin F. and Dubuisson J. (2000) J Biol Chem 275, 31428-31437;
- [13] Choo Q.L., Kuo G., Weiner A.J., Overby L.R., Bradley D.W., and Houghton M. (1989) Science 244, 359-362;
  - [14] Ciccaglione A.R., Marcantonio C., Costantino A., Equestre M., Geraci A. and Rapicetta M. (1998) Virology 250, 1-8;
- [15] Ciccaglione A.R., Marcantonio C., Equestre M.,
  15 Jones I.M. and Rapicetta M. (1998) Virus Res 55, 157165;
  - [16] Op De Beeck A., Cocquerel L., and Dubuisson J.
    (2001) J Gen Virol 82, 2589-2595;
- [17] Sharp P.M., Cowe E., Higgins D.G., Shields D.C.,
  20 Wolfe K.H., and Wright F. (1998) Nucleic Acids Res 16,
  8207-8211;
  - [18] Mullis K.B., and Faloona F.A. (1987) Methods Enzymol 155, 335-350;
- [19] Tabor S. and Richardson C.C. (1985) Proc Natl Acad Sci USA 82, 1074-1078;

[20] Guan K.L., and Dixon J.E. (1991) Anal Biochem 192, 262-267;

[21] Hakes D.J., and Dixon J.E. (1992) Anal Biochem 202, 293-298;

5 [22] Tabor S. (1990) in Current Protocols in Molecular Biology, pp. 16.12.11-16.12.11, Greene Publishing and Wiley-Interscience, New York;

[23] Schagger H. and von Jagow G. (1987) Anal Biochem 166, 368-379;

10 [24] Laemmli U.K. (1970) Nature 227, 680-685.

[25] Sambrook, Fritsch and Maniatis, Molecular cloning, A laboratory manual, second edition, Cold spring Harbor Laboratory Press, 1989.

15

#### REVENDICATIONS

- 1. Système d'expression caractérisé en ce qu'il comprend successivement, dans le sens 5'-3', une séquence nucléotidique codant pour le dipeptide Asp-Pro et une séquence nucléotidique codant pour une protéine toxique membranaire ou un domaine d'une protéine toxique membranaire.
- 2. Système d'expression selon la revendication 1, dans lequel la protéine toxique est une protéine membranaire ou un domaine d'une protéine membranaire d'une enveloppe de virus.
- 3. Système d'expression selon la revendication 2, dans lequel le virus est choisi parmi le virus de l'hépatite C, le virus du sida, un virus pathogène pour l'homme, un virus pathogène pour un mammifère.
- 4. Système d'expression selon la revendication 1, dans lequel la protéine toxique est une protéine transmembranaire ou un domaine d'une protéine transmembranaire du virus de l'hépatite C.
- 5. Système d'expression selon la revendication 1, dans lequel la protéine toxique est une protéine de séquence IDn°1 ou IDn°2 de la liste de séquences annexée.
- 306. Système d'expression selon la revendication1, dans lequel la séquence nucléotidique codant pour la

protéine toxique est choisie parmi la séquence IDn°3 et la séquence IDn°4 de la liste de séquences annexée.

- Système d'expression selon la revendication
   dans lequel la séquence nucléotidique codant pour le dipeptide Asp-Pro est gacccg.
- 8. Système d'expression selon la revendication 1, comprenant en outre, en amont de la séquence Asp10 Pro, une séquence nucléotidique codant pour une protéine soluble.
- 9. Système d'expression selon la revendication 8, dans lequel la protéine soluble est la glutathion Stransférase ou la thiorédoxine.
- 10. Système d'expression selon la revendication 1, codant pour une protéine de fusion ayant une séquence choisie dans le groupe constitué des séquences 20 IDn°46, IDn°47, IDn°48, IDn°49, IDn°50 et IDn°51 de la liste de séquences annexée.
- 11. Système d'expression selon la revendication 8, ledit système ayant une séquence choisie dans le groupe constitué des séquences IDn°34, IDn°35, IDn°36, IDn°37, IDn°38, et IDn°39 de la liste de séquences annexée.
- 12. Vecteur d'expression bactérien comprenant 30 un système d'expression selon l'une quelconque des revendications 1 à 11 cloné dans un plasmide.

- 13. Vecteur d'expression bactérien comportant un système d'expression selon l'une quelconque des revendications 1 à 9 et la séquence oligonucléotidique du plasmide pT7-7.
- 14. Vecteur d'expression bactérien constitué de la séquence IDn°44 ou IDn°45 de la liste de séquences annexée.

10

15. Vecteur d'expression bactérien comportant un système d'expression selon l'une quelconque des revendications 1 à 7 et la séquence oligonucléotidique d'un plasmide choisi parmi pGEXKT, pET32a.

15

16. Vecteur d'expression bactérien selon la revendication 15 constitué d'une séquence choisie dans le groupe constitué des séquences IDn°40, IDn°41, IDn°42 et IDn°43 de la liste de séquences annexées.

- 17. Cellule procaryote transformée par un vecteur d'expression selon l'une quelconque des revendications 12 à 16.
- 25 18. Cellule procaryote *E. coli* selon la revendication 17.
  - 19. Procédé de fabrication d'une protéine toxique par recombinaison génétique comprenant les 30 étapes suivantes :

- transformer une cellule hôte avec un système d'expression selon la revendication 1 ou avec un vecteur d'expression selon la revendication 12,
- cultiver la cellule hôte transformée dans 5 des conditions de culture telles qu'elle fabrique une protéine de fusion comprenant le dipeptide Asp-Pro suivi de la séquence peptidique de la protéine toxique à partir dudit vecteur d'expression, et
  - isoler ladite protéine de fusion.

10

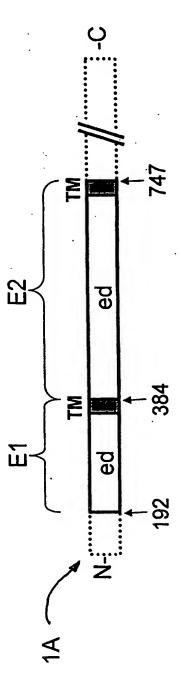
20. Procédé selon la revendication 19, comprenant en outre une étape consistant à cliver ladite protéine de fusion pour récupérer la protéine toxique.

15

- 21. Procédé selon la revendication 20, dans lequel l'étape consistant à cliver ladite protéine de fusion pour récupérer la protéine toxique est réalisée en faisant agir de l'acide formique sur la protéine de fusion.
- 22. Procédé selon la revendication 19, dans lequel la cellule hôte est *E. coli*.
- 23. Procédé selon la revendication 19, dans lequel le système d'expression code pour une protéine de fusion ayant une séquence choisie dans le groupe constitué des séquences IDn°46, IDn°47, IDn°48, IDn°49, IDn°50 et IDn°51 de la liste de séquences annexée.

- 24. Procédé selon la revendication 19, dans lequel le système d'expression a une séquence choisie dans le groupe constitué des séquences IDn°34, IDn°35, IDn°36, IDn°37, IDn°38, et IDn°39 de la liste de séquences annexée.
- 25. Procédé selon la revendication 19, dans lequel le vecteur d'expression est constitué d'une séquence choisie dans le groupe constitué des séquences 10 IDn°40, IDn°41, IDn°42, IDn°43, IDn°44 et IDn°45 de la liste de séquences annexées.
- 26. Protéine de fusion ayant une séquence peptidique choisie dans le groupe constitué des séquences IDn°46, IDn°47 IDn°48, IDn°49, IDn°50 et IDn°51 de la liste de séquences annexée.

CT/FR2003/002763



TME1: 347-MIAGAHWGVLAGIAYFSMVGNWAKVLVVLLLFAGVDA-383

TME2: 717-EYVVLLFLLLADARVCSCLWMMLLISQAEA-746

Da.	
3861	
laa,	
3.7	
347 MIAGAHWGVLAGIAYFSMVGNWAKVLVVLLLFAGVDA383	

.. GGTAACTGGGCTAAAGTTCTGGTTGTTCTGCTGCTGTTCGCTGGTGTTGACGCT ..20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 ۵ . > 90 ΛΓΛΛΓΓ × Ø 3 z 9 ပ Seq ° Z aa

C

NdeI

ClaI5' GGGAATGCCATATGATCGCTGGTG 3'

..TGGGCTAAAGTTCTGGTTGTTCTGCTGCTGTTGTTGACGCTTAGATCGATATGC 3' 131 bases .ACCCGATTTCAAGACCAACAAGACGACGACAAGCGACCACAACTGCGAATCTAGCTATACG\_5'
3'\_ACAACTGCGAATCTAGCTATACG\_5'

Clonage dans pT7-7:

OL13(+): 5'gggaatgccatatgatcgctggtg

OL14(-): 5'gcatatcgatctaagcgtcaaca

Clonage dans pGEXKT:

OL15(+): 5'ggatccatggaatacgttgttc (sans site DP)

OL17(+): 5'ggatcc**gacccg**atggaatacgttgttc (avec site DP)

OL16(-): 5'gaattcctaagcttcagcctgag

Clonage dans pET32a:

OL18(+): 5'gt**gatatc**tgatctgtctggtggtggt (s'hybride au segment 915-932 de pGEXKT)

OL16(-): 5'gaattcctaagcttcagcctgag

Clonage dans pT7-7 de (M)DP-TME1:

OL19 (+): 5'- CGCATATGGACCCGATCGCTGGTGCT -3' (Nde I souligné)

OL20 (-): 5'-GAATTCCTAAGCGTCAACACCAGC-3' (EcoR I souligné)

4	(M)	(M) $_{117}\mathrm{EYVVLLFLLLADARVCSCLWMMLLISQAEA}_{146}$	VLI	ET	LA	)AR	ZSC	LWI	P. L.	SI.	AE.	4746				3	aa	31 aa, 3546 Da.	16 1	Ja.
8	°	-	7	. m	4	5	9	7	ω	Q	10	.11		13	14	15	16 .			
	Aa	Σ	ធា	×	>	>	H	П	ĹĿı	Ы	7	· H	æ	Ω	K	æ	MEYVVLLFLLADARV			
	Seq	ATG	GAA	TAC	GTT	GTT	CTG	CTG	TTC	CTG	CIG	CIG	GCI	GAC	GCT	CGT	ATGGAATACGTTGTTCTGCTGTTCCTGCTGCTGGCTGACGCTCGTGTT			
		-		-	10			20			30		•	4	40					
	•	7	7	•	Č	ć	Č	C	č		2	ŗ	Č	Č	Č	7				
	, Z,	7	87	7	7.0	1/ 18 19 20 21 22 23 24 25 26 27 28 29 30 31	7.7	23	7.4	72	9.7	7	87	7.7	2	7			٠	
	. Aa	C S C L W M M L L I S Q A E A	S	ပ	H	3	Σ.	Σ.	H	Н	H	S	O.	Æ.	ы	Ø				
	Seq	SeqTGCTCTTGCCTGTGGATGATGCTGCTGATCTCTCAGGCTGAAGCT	CTC	rrg	CCT	GTĞ	3AT	SAT	SCT	GCT	GAT	CIC	TCA	099	TGA	AGC	H			
(		20			9	.09		Ĭ	70			80		80 90,93	<u>o</u>	6 0	m			
ر	Nd	NdeI					•			•										
S	5 - CATATGGAATACGTTGTTC 3'	ATGG	AAT!	ACG.	TTG	TTC	<del>ب</del>			•										
(+)	(+) 5 '-CATATGGAATACGTTGTTCTGCTGCTGCTGCTGGCTGACGCTCGTGTT.	ATGG/	AAT!	ACG.	TTG	TTC	ເດີ	rgr	rcc	IGC	TGC	TGG	CTG	ACG	CIC	GTG	rt			

117 bp .. TGCTCTTGCCTGTGGATGATGCTGCTGATCTCTCAGGCTGAAGCTTAAGCTT-3 .. ACGAGAACGGACACCTACTACGACGACTAGAGAGTCCGACTTCGAATTCGAA-5' 3'-GACTAGAGAGTCCGACTTCGAATTCGAA+5' Hind III

(-) 3'-GTATACCTTATGCAACAAGACGACAAGGACGACGACGACTGCGAGCACAA..

FIG. (3)

5/9

OL22(-): 5'aagcttaagcttcagcctgagagatcagcagcatcatccacaggcaagagcaaacac

Clonage dans pT7-7:

OL23(+): 5'catatggaatacgttgttc

OL24(-): 5'aagcttaagcttcagcctgagagatcag

Clonage dans pGEXKT:

OL25(+): 5'ggatccgaatacgttgttc (sans site DP)
OL27(+): 5'ggatccgacccggaatacgttgttc (avec site DP)

OL26(-): 5'gaattcttaagcttcagcctgagagatcag

Clonage dans pET32a:

OL18(+): 5'gtgatatctgatctgtctggtggtggt (s'hybride au segment 915-932 de pGEXKT)

OL26(-): 5'gaattcttaagcttcagcctgagagatcag

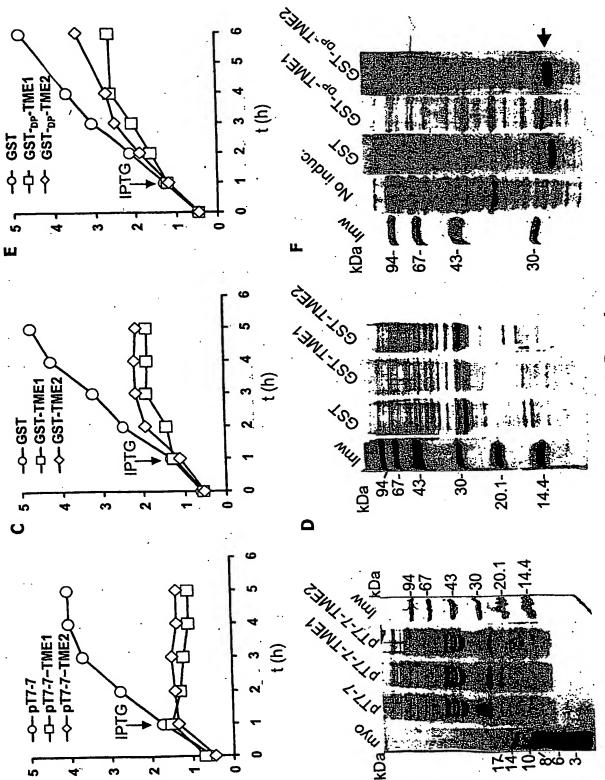
Clonage dans pT7-7 de (M)DP-TME2

OL28 (+): 5'- CGCATATGGACCCGGAATACGTTGTTC-3' (Nde I souligné)

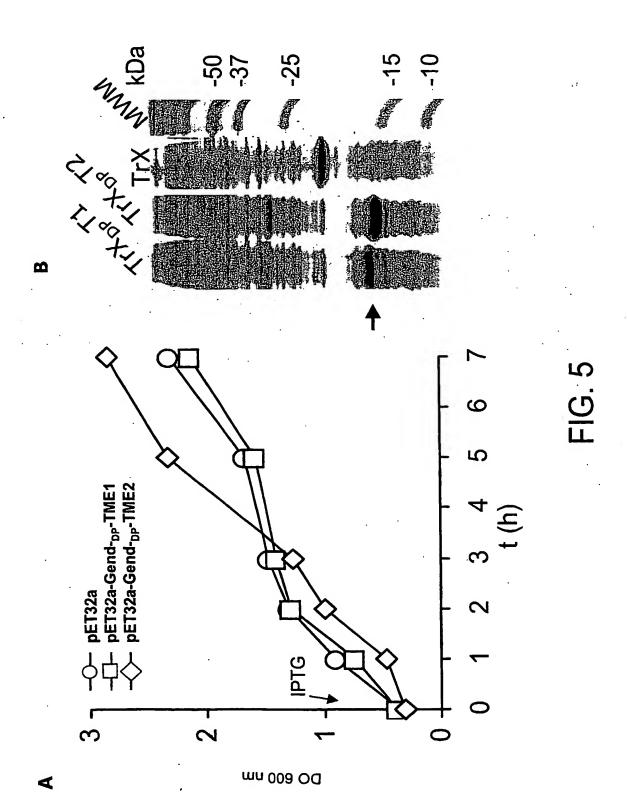
OL29 (-): 5'-CA<u>GAATTCC</u>TAAGCTTCAGCCTGAGAG-3' (EcoR I souligné)

Mn 000 OQ



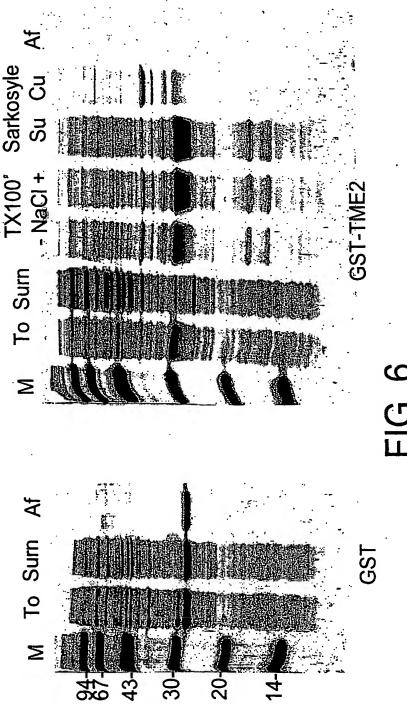


幽

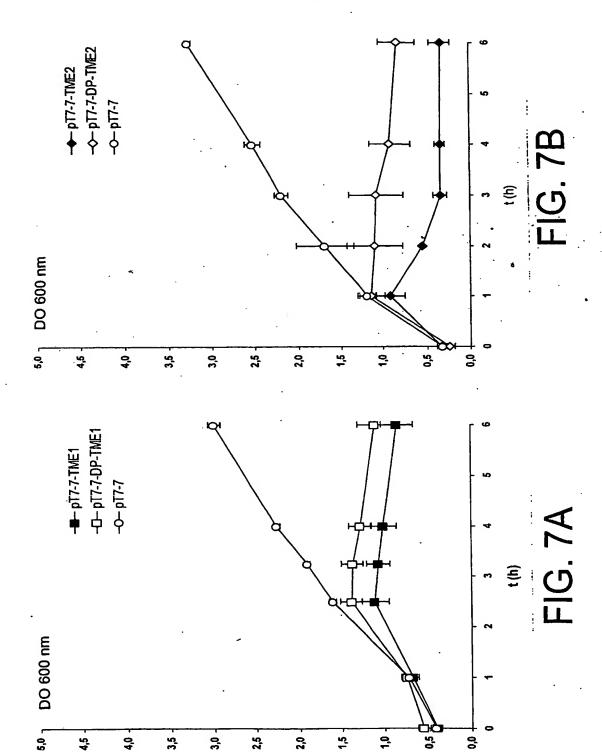


ADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLS**GGG** MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYI GGLVPRGS/PGIHRD GST

FMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLS**GGGG**L MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYIA DKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPD VPRGS/EYVVILFILLADARVCSCLWMMILISQAEA GST-TME2









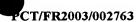
#### LISTE DE SEQUENCES

- <110> CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE .
- <120> SYSTEMES D'EXPRESSION DE PROTEINES TOXIQUES, VECTEURS ET PROCEDE DE FABRICATION DE PROTEINES TOXIQUES
- <130> B14143 EE
- <140>
- <141>
- <150> FR N°02 11676
- <151> 2002-09-20
- <160> 53
- <170> PatentIn Ver. 2.1
- <210> 1
- <211> 37
- <212> PRT
- <213> Hepatitis C virus
- <400> 1
- Met Ile Ala Gly Ala His Trp Gly Val Leu Ala Gly Ile Ala Tyr Phe

  1 5 10 15
- Ser Met Val Gly Asn Trp Ala Lys Val Leu Val Val Leu Leu Leu Phe 20 25 30
- Ala Gly Val Asp Ala 35
- <210> 2
- <211> 31
- <212> PRT
- <213> Hepatitis C virus
- <400> 2
- Met Glu Tyr Val Val Leu Leu Phe Leu Leu Leu Ala Asp Ala Arg Val 1 5 10 15
- Cys Ser Cys Leu Trp Met Met Leu Leu Ile Ser Gln Ala Glu Ala 20 25 30
- <210> 3
- <211> 111
- <212> ADN
- <213> Hepatitis C virus
- <400> 3
- atgategetg gtgeteactg gggtgttetg getggtateg ettaettete tatggttggt 60 aactgggeta aagttetggt tgttetgetg etgttegetg gtgttgaege t 111
- <210> 4
- <211> 93
- <212> ADN



<213>	Hepatitis C virus	
<400>	4	
	tacg ttgttctgct gttcctgctg ctggctgacg ctcgtgtttg ctcttgcctg atgc tgctgatctc tcaggctgaa gct	60 93
<210>	E	
<211>		
<212>		
<213>	Séquence artificielle	
<220>		•
<223>	Description de la séquence artificielle: oligonucléotide (+) d'insertion dans pT7-7	
<400>	5	
gggaat	gcca tatgatcgct ggtg	24
<210>	·	
<211><212>		
	Séquence artificielle	
<220>		٠.
	Description de la séquence artificielle: oligonucléotide (-) d'insertion dans pT7-7	
<400>	6	
gcata	togat ctaagogtca aca	23
·		
<210>	7	
<211>		
<212>		
<213>	Séquence artificielle	
<220>		
<223>	Description de la séquence artificielle: ADN sens codant de TME1 + site cla I en 3'	
<400>		
ggttg	atatg atcgctggtg ctcactgggg tgttctggct ggtatcgctt acttctctat gtaac tgggctaaag ttctggttgt tctgctgctg ttcgctggtg ttgacgctta atatg c	
<210>	. 8	
<211>		
<212>		
<213>	Séquence artificielle	
<220>		
<2235	Description de la séquence artificielle: ADN sens anticodant + site cla I en 5'	
<400>		
gcate	tegat ctaagegtea acaceagega acageageag aacaaceaga actttageee	60
	accaac catagagaag taagogatac cagecagaac accecagtga gcaccagega atggca t	120 131



<21U>	9	
<211>	74	
<212>	ADN	
<213>	Séquence artificielle	
	·	
<220>		
<223>	Description de la séquence artificielle:	
	oligonucléotide sens codant de pour la synthèse de	
	TME1	
<400>		
_	tatg atcgctggtg ctcactgggg tgttctggct ggtatcgctt acttctctat	74
aat cad	taac tggg	/ 4
<210×	10	
411>	·	
«312»		
	Séquence artificielle	
<b>&lt;220&gt;</b>		
	Description de la séquence artificielle:	
	oligonucléotide sens anticodant pour la synthèse	
	de TME1	
<b>&lt;400&gt;</b>	10	
gcatat	egat ctaagegtea acaceagega acageageag aacaaceaga actttageee	60
agttad	ccaac catagagaa	79
<210>	11	
<211>	22 .	
<212>		
<213>	Séquence artificielle	
	·	
<220>		
<223>	Description de la séquence artificielle:	
	oligonucléotide (+) d'insertion dans pGEXKT sans	
	site dp	
400	••	
<400>		22
ggate	catgg aatacgttgt tc	44
•	• •	
<210>	12	
<211>	<del></del>	
<212>		
	Séquence artificielle	
1200		
<220>		
	Description de la séquence artificielle:	
	oligonucléotide (+) d'insertion dans pGEXKT avec	
	site dp	
	-	
<400>	12	
	cgacc cgatggaata cgttgttc	28

## WO 2004/027068



<211>	23	
<212>	ADN	
<213>	Séquence artificielle	
<220>	·	
	Description de la séquence artificielle: oligonucléotide (-) d'insertion dans pGEXKT	
<400>	•	
gaatto	cctaa gcttcagcct gag	23
<210>	14	
<211>		
<212>		
<213>	Séquence artificielle	
<220>		
<223>	Description de la séquence artificielle: oligonucléotide (+) de transfert dans pET32a	
<400>	14	
gtgata	atctg atctgtctgg tggtggt	27
<210>	15	
<211>	26	
<212>	ADN	
	Séquence artificielle	
<220>		
	Description de la séquence artificielle: oligonucléotide (+) d'insertion dans pT7-7	
<400>	15	
cgcat	atgga cccgatcgct ggtgct	26
	•	
<210>	16	
<211>	24	
<212>	ADN	
<213>	Séquence artificielle	
<220>		
	Description de la séquence artificielle:	
	oligonucléotide (-) d'insertion dans pT7-7	
<400>	16	
	cctaa gcgtcaacac cagc	24
3	500aa gogoomaa ongo	
<210>	. 17	
<211>		
<211>		
	ADN Séquence artificielle	•
<220>		
	Description de la séquence artificielle:	
	oligonucléotide (+) d'insertioon dans pT7-7	
<400>	. 17	

## WO 2004/027068



	gaat acgttgttc	19
<210>		
<211><212>		
	ADN Séquence artificielle	
<220>		
	Description de la séquence artificielle: oligonucléotide (-) d'insertioon dans pT7-7	
	origonacreotide (-) d'insertioon dans pri-i	
<400>	18	
aagctt	aagc ttcagcctga gagatcag	28
·.		
<210>	19	
<211>		
<212>		
<213>	Séquence artificielle	
<220>	<b>b</b>	
<223>	Description de la séquence artificielle: ADN sens	
•	codant pour TME2 + site Nde I en 5' et Hind III en 3'	
	•·····································	
<400>	19	
	gaat acgttgttct gctgttcctg ctgctggctg acgctcgtgt ttgctcttgc	
ctgtgg	atga tgctgctgat ctctcaggct gaagcttaag ctt	103
<210>		
<211>	103	
<211> <212>	103 ADN	
<211> <212>	103	
<211> <212> <213>	103 ADN Séquence artificielle	
<211> <212> <213>	103 ADN Séquence artificielle  Description de la séquence artificielle: ADN sens	
<211> <212> <213>	103 ADN Séquence artificielle  Description de la séquence artificielle: ADN sens anticodant de TME2 + site Nde I en 3' et Hind III	
<211> <212> <213>	103 ADN Séquence artificielle  Description de la séquence artificielle: ADN sens	
<211> <212> <213> <220> <223>	103 ADN Séquence artificielle  Description de la séquence artificielle: ADN sens anticodant de TME2 + site Nde I en 3' et Hind III en 5'	
<211> <212> <213> <220> <223> <400> aagctt	103 ADN Séquence artificielle  Description de la séquence artificielle: ADN sens anticodant de TME2 + site Nde I en 3' et Hind III en 5'  20 caage tteageetga gagateagea geateateea caggeaagag caaacaegag	
<211> <212> <213> <220> <223> <400> aagctt	103 ADN Séquence artificielle  Description de la séquence artificielle: ADN sens anticodant de TME2 + site Nde I en 3' et Hind III en 5'	60 103
<211> <212> <213> <220> <223> <400> aagctt	103 ADN Séquence artificielle  Description de la séquence artificielle: ADN sens anticodant de TME2 + site Nde I en 3' et Hind III en 5'  20 caage tteageetga gagateagea geateateea caggeaagag caaacaegag geeag cageaggaac ageagaacaa egtatteeat atg	
<211> <212> <213> <220> <223> <400> aagcttcgtcac	103 ADN Séquence artificielle  Description de la séquence artificielle: ADN sens anticodant de TME2 + site Nde I en 3' et Hind III en 5'  20 caage tteageetga gagateagea geateateea caggeaagag caaacaegag gecag cageaggaac ageagaacaa egtatteeat atg	
<211> <212> <213> <220> <223> <400> aagcttcgtcac <210> <211>	103 ADN Séquence artificielle  Description de la séquence artificielle: ADN sens anticodant de TME2 + site Nde I en 3' et Hind III en 5'  20 caage tteageetga gagateagea geateateea caggeaagag caaacaegag gecag cageaggaac ageagaacaa egtatteeat atg	
<211> <212> <213> <220> <223> <400> aagcttcgtcag <210> <211> <212>	103 ADN Séquence artificielle  Description de la séquence artificielle: ADN sens anticodant de TME2 + site Nde I en 3' et Hind III en 5'  20 caage tteageetga gagateagea geateateea caggeaagag caaacaegag gecag cageaggaac ageagaacaa egtatteeat atg	
<211> <212> <213> <220> <223> <400> aagcttcgtcag <210> <211> <212> <213>	103 ADN Séquence artificielle  Description de la séquence artificielle: ADN sens anticodant de TME2 + site Nde I en 3' et Hind III en 5'  20 caage tteageetga gagateagea geateateea caggeaagag caaacaegag gecag cageaggaac ageagaacaa egtatteeat atg	
<211> <212> <213> <220> <223>  <400> aagcttcgtcag <210> <211> <212> <213> <220>	103 ADN Séquence artificielle  Description de la séquence artificielle: ADN sens anticodant de TME2 + site Nde I en 3' et Hind III en 5'  20 caage tteageetga gagateagea geateateea caggeaagag caaacaegag gecag cageaggaac ageagaacaa egtatteeat atg  21 68 ADN Séquence artificielle	
<211> <212> <213> <220> <223>  <400> aagcttcgtcag <210> <211> <212> <213> <220>	103 ADN Séquence artificielle  Description de la séquence artificielle: ADN sens anticodant de TME2 + site Nde I en 3' et Hind III en 5'  20 caage tteageetga gagateagea geateateea caggeaagag caaacaegag geag cageaggaac ageagaacaa egtatteeat atg  21 68 ADN Séquence artificielle  Description de la séquence artificielle:	
<211> <212> <213> <220> <223>  <400> aagcttcgtcag <210> <211> <212> <213> <220>	103 ADN Séquence artificielle  Description de la séquence artificielle: ADN sens anticodant de TME2 + site Nde I en 3' et Hind III en 5'  20 caage tteageetga gagateagea geateateea caggeaagag caaacaegag gecag cageaggaac ageagaacaa egtatteeat atg  21 68 ADN Séquence artificielle	
<211> <212> <213> <220> <223>  <400> aagcttcgtcag <210> <211> <212> <213> <222> <223>	ADN Séquence artificielle  Description de la séquence artificielle: ADN sens anticodant de TME2 + site Nde I en 3' et Hind III en 5'  20 caage tteageetga gagateagea geateateea caggeaagag caaacaegag gecag cageaggaac ageagaacaa egtatteeat atg  21 68 ADN Séquence artificielle  Description de la séquence artificielle: oligonucléotide sens codant (+) pour la synthèse de TME2	
<211> <212> <213> <220> <223>  <400> aagcttcgtcag <210> <211> <212> <213> <220> <223>  <400>	ADN Séquence artificielle  Description de la séquence artificielle: ADN sens anticodant de TME2 + site Nde I en 3' et Hind III en 5'  20 caage tteageetga gagateagea geateateea caggeaagag caaacaegag gecag cageaggaac ageagaacaa egtatteeat atg  21 68 ADN Séquence artificielle  Description de la séquence artificielle: oligonucléotide sens codant (+) pour la synthèse de TME2	103

## WO 2004/027068



<210>	22	
<211>	57	
<212>	ADN	
<213>	Séquence artificielle	
<220>		
	Posserintion do la sémionea antificialla.	
<2237	Description de la séquence artificielle:	
	oligonucléotide sens codant (-) pour la synthèse	
	de TME2	
<400>	22	
aagctt	aago ttoagootga gagatoagoa goatoatoca caggoaagao gaaacao	57
-		
<210>	23	
<211>		
<212>	<del></del>	
· <del>-</del>		
<213>	Séquence artificielle	
<220>		
<223>	Description de la séquence artificielle:	
	oligonucléotide (+) d'insertion dans pGEXKT sans	
	site dp	
	<del>-</del>	
<400>	23	
	cgaat acgttgttc	19
33444	-3440 40300300	13
	•	
<210>	•	
<211>		
<212>		
<213>	Séquence artificielle	
<220>	· · · · · · · · · · · · · · · · · · ·	
<223>	Description de la séquence artificielle:	
	oligonucléotide (+) d'insertion dans pGEXKT avec	
	site dp	
	arce ap	
-400-		
<400>	<del></del> -	
ggatc	cgacc cggaatacgt tgttc	25
<210>	25 .	
<211>	30	
<212>	ADN	
<213>	Séquence artificielle	
<220>		
	Description de la séquence artificielle:	
\ZZJ/	oligonucléotide (-) d'insertion dans pGEXKT avec	
	site dp	
<400>	<del></del>	
gaatt	cttaa gcttcagcct gagagatcag	30
<210>	26	
<211>		
<212>		
	Séquence artificielle	
ベムエンス	DECRETAGE OFFITTETTE	

<211> 327



```
<220>
<223> Description de la séquence artificielle:
      oligonucléotide (+) d'insertion dans pT7-7
<400> 26
cgcatatgga cccggaatac gttgttc
                                                                  27
<210> 27
<211> 27
<212> ADN
<213> Séquence artificielle
<220>
<223> Description de la séquence artificielle:
      oligonucléotide (-) d'insertion dans pT7-7
<400> 27
cagaattcct aagcttcagc ctgagag
<210> 28
<211> 15
<212> PRT
<213> Séquence artificielle
<220>
<223> Description de la séquence artificielle: fin de la
      GST suivie du site thrombine
<400> 28
Ser Asp Leu Ser Gly Gly Gly Gly Leu Val Pro Arg Gly Ser
<210> 29
<211> 717
<212> ADN
<213> Séquence artificielle
<223> Description de la séquence artificielle: ADN
      codant pour la protéine GST dans le vecteur pGEXKT
<400> 29
atgtccccta tactaggtta ttggaaaatt aagggccttg tgcaacccac tcgacttctt 60
ttggaatatc ttgaagaaaa atatgaagag catttgtatg agcgcgatga aggtgataaa 120
tggcgaaaca aaaagtttga attgggtttg gagtttccca atcttcctta ttatattgat 180
ggtgatgtta aattaacaca gtctatggcc atcatacgtt atatagctga caagcacaac 240
atgttgggtg gttgtccaaa agagcgtgca gagatttcaa tgcttgaagg agcggttttg 300
gatattagat acggtgtttc gagaattgca tatagtaaag actttgaaac tctcaaagtt 360
gattttetta geaagetace tgaaatgetg aaaatgtteg aagategttt atgteataaa 420
acatatttaa atggtgatca tgtaacccat cctgacttca tgttgtatga cgctcttgat 480
gttgttttat acatggaccc aatgtgcctg gatgcgttcc caaaattagt ttgttttaaa 540
aaacgtattg aagctatccc acaaattgat aagtacttga aatccagcaa gtatatagca 600
tggcctttgc agggctggca agccacgttt ggtggtggcg accatcctcc aaaatcggat 660
ctgtctggtg gtggtggtgg tctggttccg cgtggatccc cgggaattca tcgtgac
<210> 30
```



```
<212> ADN
<213> Séquence artificielle
<223> Description de la séquence artificielle: ADN
     codant pour la thiorédoxine dans le vecteur
     pET32a+
<400> 30
atgagegata aaattattea eetgaetgae gaeagttttg acaeggatgt aeteaaageg 60
gacggggcga tcctcgtcga tttctgggca gagtggtgcg gtccgtgcaa aatgatcgcc 120
ccgattctgg atgaaatcgc tgacgaatat cagggcaaac tgaccgttgc aaaactgaac 180
ategateaaa accetggeae tgegeegaaa tatggeatee gtggtateee gaetetgetg 240
ctgttcaaaa acggtgaagt ggcggcaacc aaagtgggtg cactgtctaa aggtcagttg 300
aaagagttcc tcgacgctaa cctggcc
<210> 31
<211> 4969
<212> ADN
<213> Séquence artificielle
<220>
<223> Description de la séquence artificielle: plasmide
     d'expression pGEXKT
<400> 31
acgttatcga ctgcacggtg caccaatgct tctggcgtca ggcagccatc ggaagctgtg 60
gtatggctgt gcaggtcgta aatcactgca taattcgtgt cgctcaaggc gcactcccgt 120
tctggataat gttttttgcg ccgacatcat aacggttctg gcaaatattc tgaaatgagc 180
tgttgacaat taatcatcgg ctcgtataat gtgtggaatt gtgagcggat aacaatttca 240
cacaggaaac agtattcatg tcccctatac taggttattg gaaaattaag ggccttgtgc 300
aacccactcg acttettttg gaatatettg aagaaaaata tgaagagcat ttgtatgage 360
gcgatgaagg tgataaatgg cgaaacaaaa agtttgaatt gggtttggag tttcccaatc 420
ttccttatta tattgatggt gatgttaaat taacacagtc tatggccatc atacgttata 480
tagctgacaa gcacaacatg ttgggtggtt gtccaaaaga gcgtgcagag atttcaatgc 540.
ttgaaggagc ggttttggat attagatacg gtgtttcgag aattgcatat agtaaagact 600
ttgaaactct caaagttgat tttcttagca agctacctga aatgctgaaa atgttcgaag 660
atogtttatg toataaaaca tatttaaatg gtgatcatgt aacccatcct gacttcatgt 720
tgtatgacgc tcttgatgtt gttttataca tggacccaat gtgcctggat gcgttcccaa 780
aattagtttg ttttaaaaaa cgtattgaag ctatcccaca aattgataag tacttgaaat 840
ccagcaagta tatagcatgg cctttgcagg gctggcaagc cacgtttggt ggtggcgacc 900
atcctccaaa atcggatctg tctggtggtg gtggtggtct ggttccgcgt ggatccccgg 960
gaattcateg tgactgactg acgatetgee tegegegttt eggtgatgae ggtgaaaace 1020
totgacacat gcagotocog gagacggtca cagottgtot gtaagoggat gcogggagca 1080
gacaagcccg tcagggcgcg tcagcgggtg ttggcgggtg tcggggcgca gccatgaccc 1140
agtcacgtag cgatagcgga gtgtataatt cttgaagacg aaagggcctc gtgatacgcc 1200
tatttttata ggttaatgtc atgataataa tggtttctta gacgtcaggt ggcacttttc 1260
ggggaaatgt gcgcggaacc cctatttgtt tatttttcta aatacattca aatatgtatc 1320
cgctcatgag acaataaccc tgataaatgc ttcaataata ttgaaaaagg aagagtatga 1380
gtattcaaca tttccgtgtc gcccttattc ccttttttgc ggcattttgc cttcctgttt 1440
ttgctcaccc agaaacgctg gtgaaagtaa aagatgctga agatcagttg ggtgcacgag 1500
tgggttacat cgaactggat ctcaacagcg gtaagatcct tgagagtttt cgccccgaag 1560
aacgttttcc aatgatgagc acttttaaag ttctgctatg tggcgcggta ttatcccgtg 1620
ttgacgccgg gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat gacttggttg 1680
agtactcacc agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca 1740
gtgctgccat aaccatgagt gataacactg cggccaactt acttctgaca acgatcggag 1800
gaccgaagga gctaaccgct tttttgcaca acatggggga tcatgtaact cgccttgatc 1860
gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg 1920
cagcaatggc aacaacgttg cgcaaactat taactggcga actacttact ctagcttccc 1980
ggcaacaatt aatagactgg atggaggcgg ataaagttgc aggaccactt ctgcgctcgg 2040
```



```
ecctteegge tggetggttt attgetgata aatetggage eggtgagegt gggtetegeg 2100
gtatcattgc agcactgggg ccagatggta agccctcccg tatcgtagtt atctacacga 2160
cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac 2220
tgattaagca ttggtaactg tcagaccaag tttactcata tatactttag attgatttaa 2280
aacttcattt ttaatttaaa aggatctagg tgaagatcct ttttgataat ctcatgacca 2340
aaatccctta acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag 2400
gatcttcttg agatcctttt tttctgcgcg taatctgctg cttgcaaaca aaaaaaccac 2460
cgctaccage ggtggtttgt ttgccggate aagagetace aactetttt ccgaaggtaa 2520
etggetteag cagagegeag ataccaaata etgteettet agtgtageeg tagttaggee 2580
sesacttcaa gaactctgta gcaccgccta catacctcgc tctgctaatc ctgttaccag 2640
egctgctgc cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac 2700
eggataagge geageggteg ggetgaaegg ggggttegtg cacacagece agettggage 2760
gaacgaccta caccgaactg agatacctac agcgtgagct atgagaaagc gccacgcttc 2820
ccgaagggag aaaggcggac aggtatccgg taagcggcag ggtcggaaca ggagagcgca 2880
Magggaget tecaggggga aacgeetggt atetttatag teetgteggg tttegeeace 2940
tetgaettga gegtegattt ttgtgatget egteaggggg geggageeta tggaaaaacg 3000
reageaacge ggeettttta eggtteetgg cettttgetg geettttget cacatgttet 3060
* coctgogtt atcocctgat totgtggata acceptattac cecetttegae teaectgata 3120
 egetegeeg cageegaaeg acegagegea gegagteagt gagegaggaa geggaagage 3180
ectgatgeg gtattttete ettaegeate tgtgeggtat tteacacege ataaatteeg 3240
*caccatcga atggtgcaaa acctttcgcg gtatggcatg atagcgcccg gaagagagtc 3300
aattcagggt ggtgaatgtg aaaccagtaa cgttatacga tgtcgcagag tatgccggtg 3360
totottatca gaccgtttcc cgcgtggtga accaggccag ccacgtttct gcgaaaacgc 3420
gggaaaaagt ggaagcggcg atggcggagc tgaattacat tcccaaccgc gtggcacaac 3480
aactggeggg caaacagteg tigetgatig gegtigeeac ciecagteig geecigeaeg 3540
cgccgtcgca aattgtcgcg gcgattaaat ctcgcgccga tcaactgggt gccagcgtgg 3600
tggtgtcgat ggtagaacga agcggcgtcg aagcctgtaa agcggcggtg cacaatcttc 3660
tcgcgcaacg cgtcagtggg ctgatcatta actatccgct ggatgaccag gatgccattg 3720
ctgtggaagc tgcctgcact aatgttccgg cgttatttct tgatgtctct gaccagacac 3780
ccatcaacag tattattttc tcccatgaag acggtacgcg actgggcgtg gagcatctgg 3840
tegeattggg teaceageaa ategegetgt tagegggeee attaagttet gteteggege 3900
gtctgcgtct ggctggctgg cataaatatc tcactcgcaa tcaaattcag ccgatagcgg 3960
aacgggaagg cgactggagt gccatgtccg gttttcaaca aaccatgcaa atgctgaatg 4020
aggicatingt teccacting atgettigting coalangatea gatggenering generating 4080
gegecattae egagteeggg etgegegttg gtgeggatat eteggtagtg ggataegaeg 4140
ataccgaaga cageteatgt tatatecege egttaaceae cateaaacag gattttegee 4200
tgctggggca aaccagcgtg gaccgcttgc tgcaactctc tcagggccag gcggtgaagg 4260
gcaatcagct gttgcccgtc tcactggtga aaagaaaaac caccctggcg cccaatacgc 4320
aaaccgcctc tccccgcgcg ttggccgatt cattaatgca gctggcacga caggtttccc 4380
gactggaaag cgggcagtga gcgcaacgca attaatgtga gttagctcac tcattaggca 4440
ccccaggett tacactttat getteegget egtatgttgt gtggaattgt gageggataa 4500
caatttcaca caggaaacag ctatgaccat gattacggat tcactggccg tcgttttaca 4560
acgtcgtgac tgggaaaacc ctggcgttac ccaacttaat cgccttgcag cacatccccc 4620
tttegecage tggegtaata gegaagagge cegeacegat egecetteee aacagttgeg 4680
cageetgaat ggegaatgge getttgeetg gttteeggea eeagaagegg tgeeggaaag 4740
ctggctggag tgcgatcttc ctgaggccga tactgtcgtc gtcccctcaa actggcagat 4800
gcacggttac gatgcgccca tctacaccaa cgtaacctat cccattacgg tcaatccgcc 4860
gtttgttccc acggagaatc cgacgggttg ttactcgctc acatttaatg ttgatgaaag 4920
ctggctacag gaaggccaga cgcgaattat ttttgatggc gttggaatt
```

```
<210> 32
<211> 11800
<212> ADN
<213> Séquence artificielle
<220>
<223> Description de la séquence artificielle: plasmide d'expression pET32a+
```

<400> 32





atccggatat agttcctcct ttcagcaaaa aacccctcaa gacccgttta gaggccccaa 60 ggggttatgc tagttattgc tcagcggtgg cagcagccaa taggcctata tcaaggagga 120 aagtcgtttt ttggggagtt ctgggcaaat ctccggggtt ccccaatacg atcaataacg 180 agtegeeace gtegteggtt eteagettee tttegggett tgttageage eggateteag 240 tggtggtggt ggtggtgctc gagtgcggcc gcaagcttgt cgacggagct cgaattcgga 300 gagtogaagg aaagcccgaa acaatcgtcg gcctagagtc accaccacca ccaccacgag 360 ctcacgccgg cgttcgaaca gctgcctcga gcttaagcct tccgatatca gccatggcct 420 tgtcgtcgtc gtcggtaccc agatctgggc tgtccatgtg ctggcgttcg aatttagcag 480 cagcggtttc tttcatacca aggctatagt cggtaccgga acagcagcag cagccatggg 540 totagaccog acaggtacac gaccgcaagc ttaaatcgtc gtcgccaaag aaagtatggt 600 gaaccgcgtg gcaccagacc agaagaatga tgatgatgat ggtgcatatg gccagaacca 660 gaaccggcca ggttagcgtc gaggaactct ttcaactgac cttggcgcac cgtggtctgg 720 tettettaet actactaeta ceaegtatae eggtettggt ettggeeggt ceaategeag 780 ctccttgaga aagttgactg ctttagacag tgcacccact ttggttgccg ccacttcacc 840 gtttttgaac agcagcagag tcgggatacc acggatgcca tatttcggcg cagtgccagg 900 gaaatctgtc acgtgggtga aaccaacggc ggtgaagtgg caaaaacttg tcgtcgtctc 960 agecetatgg tgeetaeggt ataaageege gteaeggtee gttttgateg atgtteagtt 1020 ttgcaacggt cagtttgccc tgatattcgt cagcgatttc atccagaatc ggggcgatca 1080 ttttgcacgg accgcaccac caaaactagc tacaagtcaa aacgttgcca gtcaaacggg 1140 actataagca gtcgctaaag taggtcttag ccccgctagt aaaacgtgcc tggcgtggtg 1200 tetgeceaga aategaegag gategeeeg teegetttga gtacateegt gteaaaactg 1260 tegteagtea ggtgaataat tttategete atatgtatat agaegggtet ttagetgete 1320 ctagcggggc aggcgaaact catgtaggca cagttttgac agcagtcagt ccacttatta 1380 aaatagcgag tatacatata ctccttctta aagttaaaca aaattatttc tagaggggaa 1440 ttgttatccg ctcacaattc ccctatagtg agtcgtatta atttcgcggg atcgagatcg 1500 gaggaagaat ttcaatttgt tttaataaag atctcccctt aacaataggc gagtgttaag 1560 gggatatcac tcagcataat taaagcgccc tagctctagc atctcgatcc tctacgccgg 1620 acgcategtg geeggeatea ceggegeeae aggtgeggtt getggegeet atategeega 1680 catcaccgat ggggaagatc tagagctagg agatgcggcc tgcgtagcac cggccgtagt 1740 ggccgcggtg tccacgccaa cgaccgcgga tatagcggct gtagtggcta ccccttctag 1800 gggctcgcca cttcgggctc atgagcgctt gtttcggcgt gggtatggtg gcaggccccg 1860 tggccggggg actgttgggc gccatctcct tgcatgcacc cccgagcggt gaagcccgag 1920 tactcgcgaa caaagccgca cccataccac cgtccggggc accggcccc tgacaacccg 1980 cggtagagga acgtacgtgg attccttgcg gcggcggtgc tcaacggcct caacctacta 2040 etgggetget teetaatgea ggagtegeat aagggagage gtegagatee eggaeaceat 2100 taaggaacgc cgccgccacg agttgccgga gttggatgat gacccgacga aggattacgt 2160 cctcagcgta ttccctctcg cagctctagg gcctgtggta cgaatggcgc aaaacctttc 2220 geggtatgge atgatagege eeggaagaga gteaatteag ggtggtgaat gtgaaaceag 2280 taacgttata cgatgtcgca gcttaccgcg ttttggaaag cgccataccg tactatcgcg 2340 ggccttctct cagttaagtc ccaccactta cactttggtc attgcaatat gctacagcgt 2400 gagtatgccg gtgtctctta tcagaccgtt tcccgcgtgg tgaaccaggc cagccacgtt 2460 tctgcgaaaa cgcgggaaaa agtggaagcg gcgatggcgg ctcatacggc cacagagaat 2520 agtetggcaa agggegeace acttggteeg gteggtgeaa agaegetttt gegeeetttt 2580 tcaccttcgc cgctaccgcc agctgaatta cattcccaac cgcgtggcac aacaactggc 2640 gggcaaacag tcgttgctga ttggcgttgc cacctccagt ctggccctgc acgcgccqtc 2700 tcgacttaat gtaagggttg gcgcaccgtg ttgttgaccg cccgtttgtc agcaacgact 2760 aaccgcaacg gtggaggtca gaccgggacg tgcgcggcag gcaaattgtc gcggcgatta 2820 aatctcgcgc cgatcaactg ggtgccagcg tggtggtgtc gatggtagaa cgaagcggcg 2880 tcgaagcctg taaagcggcg cgtttaacag cgccgctaat ttagagcgcg gctagttgac 2940 ccacggtcgc accaccacag ctaccatctt gcttcgccgc agcttcggac atttcgccgc 3000 gtgcacaatc ttctcgcgca acgcgtcagt gggctgatca ttaactatcc gctggatgac 3060 caggatgcca ttgctgtgga agctgcctgc actaatgttc cacgtgttag aagagcgcgt 3120 tgcgcagtca cccgactagt aattgatagg cgacctactg gtcctacggt aacgacacct 3180 tcgacggacg tgattacaag cggcgttatt tcttgatgtc tctgaccaga cacccatcaa 3240 cagtattatt ttctcccatg aagacggtac gcgactgggc gtggagcatc tggtcgcatt 3300 gccgcaataa agaactacag agactggtct gtgggtagtt gtcataataa aagagggtac 3360 ttctgccatg cgctgacccg cacctcgtag accagcgtaa gggtcaccag caaatcgcgc 3420 tgttagcggg cccattaagt tetgtetegg egegtetgeg tetggetgge tggcataaat 3480 ateteaeteg caateaaatt eecagtggte gtttagegeg acaategeee gggtaattea 3540 agacagagee gegeagaege agacegaeeg acegtattta tagagtgage gttagtttaa 3600 cageegatag eggaaeggga aggegaetgg agtgecatgt eeggttttea acaaaceatg 3660



caaatgctga atgagggcat cgttcccact gcgatgctgg gtcggctatc gccttgccct 3720 teegetgace teaeggtaca ggccaaaagt tgtttggtac gtttacgact tacteeggta 3780 gcaagggtga cgctacgacc ttgccaacga tcagatggcg ctgggcgcaa tgcgcgccat 3840 taccgagtcc gggctgcgcg ttggtgcgga catctcggta gtgggatacg acgataccga 3900 aacggttgct agtctaccgc gacccgcgtt acgcgcggta atggctcagg cccgacgcgc 3960 aaccacgcct gtagagccat caccctatgc tgctatggct agacagctca tgttatatcc 4020 egecgttaac caccatcaaa caggattttc geetgetggg geaaaccage gtggaccget 4080 tgctgcaact ctctcagggc tctgtcgagt acaatatagg gcggcaattg gtggtagttt 4140 gtcctaaaag cggacgaccc cgtttggtcg cacctggcga acgacgttga gagagtcccg 4200 caggoggtga agggcaatca getgttgccc gtctcactgg tgaaaagaaa aaccaccctg 4260 gegeceaata egeaaacege eteteceege gegttggeeg gteegeeact teeegttagt 4320 cgacaacggg cagagtgacc acttttcttt ttggtgggac cgcgggttat gcgtttggcg 4380 gagaggggcg cgcaaccggc attcattaat gcagctggca cgacaggttt cccgactgqa 4440 aagcgggcag tgagcgcaac gcaattaatg taagttagct cactcattag gcaccgggat 4500 taagtaatta cgtcgaccgt gctgtccaaa gggctgacct ttcgcccgtc actcgcgttg 4560 cgttaattac attcaatcga gtgagtaatc cgtggcccta ctcgaccgat gcccttgaga 4620 gccttcaacc cagtcagctc cttccggtgg gcgcggggca tgactatcgt cgccgcactt 4680 atgactgtct tetttateat gagetggeta egggaactet eggaagttgg gteagtegag 4740 gaaggccacc cgcgccccgt actgatagca gcggcgtgaa tactgacaga agaaatagta 4800 gcaactegta ggacaggtge eggeageget etgggteatt tteggegagg acegettteg 4860 ctggagcgcg acgatgatcg gcctgtcgct tgcggtattc cgttgagcat cctgtccacg 4920 gccgtcgcga gacccagtaa aagccgctcc tggcgaaagc gacctcgcgc tgctactagc 4980 cggacagcga acgccataag ggaatcttgc acgccctcgc tcaagccttc gtcactggtc 5040 cegecaceaa aegtttegge gagaageagg ceattatege eggeatggeg geeceaeggg 5100 cettagaacg tgcgggagcg agttcggaag cagtgaccag ggcggtggtt tgcaaagccg 5160 ctettegtee ggtaatageg geegtacege eggggtgeee tgegeatgat egtgeteetg 5220 tcgttgagga cccggctagg ctggcggggt tgccttactg gttagcagaa tgaatcaccg 5280 atacgcgagc gaacgtgaag acgcgtacta gcacgaggac agcaactcct gggccgatcc 5340 gaccgcccca acggaatgac caatcgtctt acttagtggc tatgcgctcg cttgcacttc 5400 cgactgctgc tgcaaaacgt ctgcgacctg agcaacaaca tgaatggtct tcggtttccg 5460 tgtttcgtaa agtctggaaa cgcggaagtc agcgccctgc gctgacgacg acgttttgca 5520 gacgetggae tegttgttgt acttaccaga agecaaagge acaaageatt teagacettt 5580 gegeetteag tegegggaeg accattatgt teeggatetg categeagga tgetgetgge 5640 taccetgtgg aacacetaca tetgtattaa egaagegetg geattgacee tgagtgattt 5700 tggtaataca aggcctagac gtagcgtcct acgacgaccg atgggacacc ttgtggatgt 5760 agacataatt gettegegae egtaactggg acteactaaa ttetetggte eegeegeate 5820 cataccycca gttgtttacc ctcacaacgt tccagtaacc gggcatgttc atcatcagta 5880 accegtateg tgageateet aagagaeeag ggeggegtag gtatggeggt caacaaatgg 5940 gagtgttgca aggtcattgg cccgtacaag tagtagtcat tgggcatagc actcgtagga 6000 ctctcgtttc atcggtatca ttacccccat gaacagaaat cccccttaca cggaggcatc 6060 agtgaccaaa caggaaaaaa ccgcccttaa catggcccgc gagagcaaag tagccatagt 6120 aatgggggta cttgtcttta gggggaatgt gcctccgtag tcactggttt gtccttttt 6180 ggcgggaatt gtaccgggcg tttatcagaa gccagacatt aacgcttctg gagaaactca 6240 acgagetgga egeggatgaa caggeagaca tetgtgaate getteaegae caegetgatg 6300 aaatagtett eggtetgtaa ttgegaagae etetttgagt tgetegaeet gegeetaett 6360 gtoogtotgt agacacttag cgaagtgotg gtgcgactac agctttaccg cagetgcctc 6420 gegegttteg gtgatgaegg tgaaaacete tgacacatge ageteeegga gaeggteaca 6480 gettgtetgt aageggatge tegaaatgge gtegaeggag egegeaaage eactaetgee 6540 acttttggag actgtgtacg tcgagggcct ctgccagtgt cgaacagaca ttcgcctacg 6600 cgggagcaga caagcccgtc agggcgcgtc agcgggtgtt ggcgggtgtc ggggcgcagc 6660 catgacccag tcacgtagcg atageggagt gtatactggc gccctegtct gttcgggcag 6720 tecegegeag tegeceacaa eegeceacag eecegegteg gtactgggte agtgeatege 6780 tatogoctca catatgacog ttaactatgo ggcatcagag cagattgtac tgagagtgca 6840 ccatatatgc ggtgtgaaat accgcacaga tgcgtaagga gaaaataccg catcaggcgc 6900 aattgatacg ccgtagtctc gtctaacatg actctcacgt ggtatatacg ccacacttta 6960 tggcgtgtct acgcattcct cttttatggc gtagtccgcg tcttccgctt cctcgctcac 7020 tgactcgctg cgctcggtcg ttcggctgcg gcgagcggta tcagctcact caaaggcggt 7080 aatacggtta tecacagaat agaaggegaa ggagegagtg aetgagegae gegageeage 7140 aagccgacgc cgctcgccat agtcgagtga gtttccgcca ttatgccaat aggtgtctta 7200 caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta 7260 aaaaggccgc gttgctggcg tttttccata ggctccgccc gtcccctatt gcgtcctttc 7320



ttgtacactc gttttccggt cgttttccgg tccttggcat ttttccggcg caacgaccgc 7380 aaaaaggtat ccgaggcggg ccctgacgag catcacaaaa atcgacgctc aagtcagagg 7440 tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag ctccctcgtg 7500 gggactgete gtagtgtttt tagetgegag tteagtetee acegetttgg getgteetga 7560 tatttctatg gtccgcaaag ggggaccttc gagggagcac cgctctcctg ttccgaccct 7620 geogettace ggatacetgt ecgeettet ecetteggga agegtggege ttteteatag 7680 ctcacgctgt aggtatctca gcgagaggac aaggctggga cggcgaatgg cctatggaca 7740 ggcggaaaga gggaagccct tcgcaccgcg aaagagtatc gagtgcgaca tccatagagt 7800 gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaacccccc gttcagcccg 7860 accgctgcgc cttatccggt aactatcgtc ttgagtccaa caagccacat ccagcaagcg 7920 aggttcgacc cgacacacgt gcttgggggg caagtcgggc tggcgacgcg gaataggcca 7980 ttgatagcag aactcaggtt cccggtaaga cacgacttat cgccactggc agcagccact 8040 ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagttctt gaagtggtgg 8100 gggccattet gtgctgaata geggtgaeeg tegteggtga ceattgteet aategteteg 8160 ctccatacat ccgccacgat gtctcaagaa cttcaccacc cctaactacg gctacactag 8220 aaggacagta titggtatot gegetetget gaagecagtt acetteggaa aaagagttgg 8280 tagetettga teeggeaaac ggattgatge egatgtgate tteetgteat aaaccataga 8340 cgcgagacga cttcggtcaa tggaagcctt tttctcaacc atcgagaact aggccgtttg 8400 aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg cgcagaaaaa 8460 aaggatetea agaagateet ttgatetttt etaeggggte tttggtggeg accategeea 8520 ccaaaaaaac aaacgttcgt cgtctaatgc gcgtcttttt ttcctagagt tcttctagga 8580 aactagaaaa gatgccccag tgacgctcag tggaacgaaa actcacgtta agggattttg 8640 gtcatgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa atgaagtttt 8700 actgcgagtc accttgcttt tgagtgcaat tecctaaaac cagtactcta atagtttttc 8760 ctagaagtgg atctaggaaa atttaatttt tacttcaaaa aaatcaatct aaagtatata 8820 tgagtaaact tggtctgaca gttaccaatg cttaatcagt gaggcaccta tctcagcgat 8880 ctgtctattt cgttcatcca tttagttaga tttcatatat actcatttga accagactgt 8940 caatggttac gaattagtca ctccgtggat agagtcgcta gacagataaa gcaagtaggt 9000 tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc 9060 ccagtgctgc aatgataccg cgagacccac gctcaccggc atcaacggac tgaggggcag 9120 cacatctatt gatgctatgc cctcccgaat ggtagaccgg ggtcacgacg ttactatggc 9180 gctctgggtg cgagtggccg tccagattta tcagcaataa accagccagc cggaagggcc 9240 gagcgcagaa gtggtcctgc aactttatcc gcctccatcc agtctattaa ttgttgccgg 9300 aggictaaat agicgitatt iggicggicg gccticccgg cicgcgicti caccaggacg 9360 ttgaaatagg cggaggtagg tcagataatt aacaacggcc gaagctagag taagtagttc 9420 gccagttaat agtttgcgca acgttgttgc cattgctgca ggcatcgtgg tgtcacgctc 9480 gtcgtttggt atggcttcat cttcgatctc attcatcaag cggtcaatta tcaaacgcgt 9540 tgcaacaacg gtaacgacgt ccgtagcacc acagtgcgag cagcaaacca taccgaagta 9600 tcagctccgg ttcccaacga tcaaggcgag ttacatgatc ccccatgttg tgcaaaaaag 9660 cggttagctc cttcggtcct ccgatcgttg tcagaagtaa agtcgaggcc aagggttgct 9720 agttccgctc aatgtactag ggggtacaac acgttttttc gccaatcgag gaagccagga 9780 ggctagcaac agtcttcatt gttggccgca gtgttatcac tcatggttat ggcagcactg 9840 cataattete ttaetgteat gecateegta agatgetttt etgtgaetgg tgagtaetea 9900 caaccggcgt cacaatagtg agtaccaata ccgtcgtgac gtattaagag aatgacagta 9960 cggtaggcat tctacgaaaa gacactgacc actcatgagt accaagtcat tctgagaata 10020 gtgtatgcgg cgaccgagtt gctcttgccc ggcgtcaata cgggataata ccgcgccaca 10080 tagcagaact ttaaaagtgc tggttcagta agactcttat cacatacgcc gctggctcaa 10140 cgagaacggg ccgcagttat gccctattat ggcgcggtgt atcgtcttga aattttcacg 10200 tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat 10260 ccagttcgat gtaacccact cgtgcaccca actgatcttc agtagtaacc ttttgcaaga 10320 agccccgctt ttgagagttc ctagaatggc gacaactcta ggtcaagcta cattgggtga 10380 gcacgtgggt tgactagaag agcatctttt actttcacca gcgtttctgg gtgagcaaaa 10440 acaggaaggc aaaatgccgc aaaaaaggga ataagggcga cacggaaatg ttgaatactc 10500 togtagaaaa tgaaagtggt ogcaaagaco cactogtttt tgtocttoog ttttacggog 10560 ttttttccct tattcccgct gtgcctttac aacttatgag atactcttcc tttttcaata 10620 ttattgaagc atttatcagg gttattgtct catgagcgga tacatatttg aatgtattta 10680 gaaaaataaa caaatagggg tatgagaagg aaaaagttat aataacttcg taaatagtcc 10740 caataacaga gtactcgcct atgtataaac ttacataaat ctttttattt gtttatcccc 10800 ttccgcgcac atttccccga aaagtgccac ctgaaattgt aaacgttaat attttgttaa 10860 aattcgcgtt aaatttttgt taaatcagct catttttaa aaggcgcgtg taaaggggct 10920 tttcacggtg gactttaaca tttgcaatta taaaacaatt ttaagcgcaa tttaaaaaca 10980



```
atttagtcga gtaaaaaatt ccaataggcc gaaatcggca aaatccctta taaatcaaaa 11040
gaatagaccg agatagggtt gagtgttgtt ccagtttgga acaagagtcc actattaaag 11100
ggttatccgg ctttagccgt tttagggaat atttagtttt cttatctggc tctatcccaa 11160
ctcacaacaa ggtcaaacct tgttctcagg tgataatttc aacgtggact ccaacgtcaa 11220
agggcgaaaa accgtctatc agggcgatgg cccactacgt gaaccatcac cctaatcaag 11280
ttttttgggg tcgaggtgcc ttgcacctga ggttgcagtt tcccgctttt tggcagatag 11340
tecegetace gggtgatgea ettggtagtg ggattagtte aaaaaacece agetecacgg 11400
gtaaagcact aaatcggaac cctaaaggga gcccccgatt tagagcttga cggggaaagc 11460
cggcgaacgt ggcgagaaag gaagggaaga aagcgaaagg catttcgtga tttagccttg 11520
ggatttccct cgggggctaa atctcgaact gcccctttcg gccgcttgca ccgctctttc 11580
cttcccttct ttcgctttcc agcgggcgct agggcgctgg caagtgtagc ggtcacgctg 11640
egegtaacca ccacaccege egegettaat gegeegetae agggegegte ccattegeea 11700
tegecegega tecegegace gtteacateg ceagtgegae gegeattggt ggtgtgggeg 11760
gcgcgaatta cgcggcgatg tcccgcgcag ggtaagcggt
<210> 33
<211> 2504
<212> ADN
<213> Séquence artificielle
<220>
<223> Description de la séquence artificielle: plasmide
      d'expression pT7-7
<400> 33
aatteteatg tttgacaget tateategat gataagettg ggetgeaggt cgactetaga 60
ggatccccgg gcgcgaattc tagccatatg tatatctcct tcttaaagtt aaacaaaatt 120
atttctagag ggaaaccgtt gtggtctccc tatagtgagt cgtattaatt tcgaagtcta 180
tcagaagttc gaatcgctgg gcctcgcgcg tttcggtgat gacggtgaaa acctctgaca 240
catgcagete eeggagaegg teacagettg tetgtaageg gatgeeggga geagacaage 300
ccgtcagggc gcgtcagcgg gtgttggcgg gtgtcggggc gcagccatga cccagtcacg 360
tagcgatagc ggagtgtata tactggctta actatgcggc atcagagcag attgtactga 420
gagtgcacca taggaagatc ttccggaaga tcttcctatg cggtgtgaaa taccgcacag 480
atgogtaagg agaaaatacc gcatcaggcg ctcttccgct tcctcgctca ctgactcgct 540
gegeteggte gtteggetge ggegageggt ateageteae teaaaggegg taataeggtt 600
atccacagaa tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc 660
caggaaccgt aaaaaggccg cgttgctggc gtttttccat aggctccgcc cccctgacga 720
gcatcacaaa aatcgacgct caagtcagag gtggcgaaac ccgacaggac tataaagata 780
ccaggcgttt ccccctggaa gctccctcgt gcgctctcct gttccgaccc tgccgcttac 840
cggatacctg tccgcctttc tcccttcggg aagcgtggcg ctttctcaat gctcacgctg 900
taggtatete agtteggtgt aggtegtteg etecaagetg ggetgtgtge aegaacece 960
cgttcagccc gaccgctgcg ccttatccgg taactatcgt cttgagtcca acccggtaag 1020
acacgaetta tegecaetgg cageageeac tggtaacagg attageagag egaggtatgt 1080
aggeggtget acagagttet tgaagtggtg geetaaetae ggetaeaeta gaaggaeagt 1140
atttggtate tgegetetge tgaageeagt tacettegga aaaagagttg gtagetettq 1200
atceggeaaa caaaccaccg etggtagegg tggttttttt gtttgeaage agcagattae 1260
gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt tctacggggt ctgacgctca 1320
gtggaacgaa aactcacgtt aagggatttt ggtcatgaga ttatcaaaaa ggatcttcac 1380
ctagateett ttaattettg aagaegaaag ggeetegtga taegeetatt tttataggtt 1440
aatgtcatga taataatggt ttcttagacg tcaggtggca cttttcgggg aaatgtgcgc 1500
ggaaccccta tttgtttatt tttctaaata cattcaaata tgtatccgct catgagacaa 1560
taaccctgat aaatgcttca ataatattga aaaaggaaga gtatgagtat tcaacatttc 1620.
cgtgtcgccc ttattccctt ttttgcggca ttttgccttc ctgtttttgc tcacccagaa 1680
acgctggtga aagtaaaaga tgctgaagat cagttgggtg cacgagtggg ttacatcgaa 1740
ctggatctca acagcggtaa gatccttgag agttttcgcc ccgaagaacg ttttccaatg 1800
atgagcactt ttaaagttct gctatgtggc gcggtattat cccgtgttga cgccggqcaa 1860
gagcaactcg gtcgccgcat acactattct cagaatgact tggttgagta ctcaccagtc 1920
acagaaaagc atcttacgga tggcatgaca gtaagagaat tatgcagtgc tgccataacc 1980
atgagtgata acactgegge caacttactt etgacaacga teggaggace gaaggageta 2040
accgcttttt tgcacaacat gggggatcat gtaactcgcc ttgatcgttg ggaaccggag 2100
```

<212> ADN



```
ctgaatgaag ccataccaaa cgacgagcgt gacaccacga tgcctgtagc aatggcaaca 2160
acgttgcgca aactattaac tggcgaacta cttactctag cttcccggca acaattaata 2220
gactggatgg aggcggataa agttgcagga ccacttctgc gctcggccct tccggctggc 2280
tggtttattg ctgataaatc tggagccggt gagcgtgggt ctcgcggtat cattgcagca 2340
ctggggccag atggtaagcc ctcccgtatc gtagttatct acacgacggg gagtcaggca 2400
actatggatg aacgaaatag acagatcgct gagataggtg cctcactgat taagcattgg 2460
taactgtcag accaagttta ctcatatata ctttagattg attt
<210> 34
<211> 813
<212> ADN
<213> Séquence artificielle
<220>
<223> Description de la séquence artificielle: système
      d'expression codant pour la protéine de fusion
<400> 34
atgtccccta tactaggtta ttggaaaatt aagggccttg tgcaacccac tcgacttctt 60
ttggaatatc ttgaagaaaa atatgaagag catttgtatg agcgcgatga aggtgataaa 120
tggcgaaaca aaaagtttga attgggtttg gagtttccca atcttcctta ttatattgat 180
ggtgatgtta aattaacaca gtctatggcc atcatacgtt atatagctga caagcacaac 240
atgttgggtg gttgtccaaa agagcgtgca gagatttcaa tgcttgaagg agcggttttg 300
gatattagat acggtgtttc gagaattgca tatagtaaag actttgaaac tctcaaagtt 360
gattttctta gcaagctacc tgaaatgctg aaaatgttcg aagatcgttt atgtcataaa 420
acatatttaa atggtgatca tgtaacccat cctgacttca tgttgtatga cgctcttgat 480
gttgttttat acatggaccc aatgtgcctg gatgcgttcc caaaattagt ttgttttaaa 540
aaacgtattg aagctatccc acaaattgat aagtacttga aatccagcaa gtatatagca 600
tggcctttgc agggctggca agccacgttt ggtggtggcg accatcctcc aaaatcggat 660
ctgtctggtg gtggtggtgg tctggttccg cgtggatccg acccgatcgc tggtgctcac 720
tggggtgttc tggctggtat cgcttacttc tctatggttg gtaactgggc taaagttctg 780
gttgttctgc tgctgttcgc tggtgttgac gct
                                                                   813
<210> 35
<211> 513
<212> ADN
<213> Séquence artificielle
<220>
<223> Description de la séquence artificielle: système
      d'expression codant pour la protéine de fusion
      TrX-DP-TME1
<400> 35
atgagcgata aaattattca cctgactgac gacagttttg acacggatgt actcaaagcg 60
gacggggcga tcctcgtcga tttctgggca gagtggtgcg gtccgtgcaa aatgatcgcc 120
ccgattctgg atgaaatcgc tgacgaatat cagggcaaac tgaccgttgc aaaactgaac 180
ategateaaa accetggeae tgegeegaaa tatggeatee gtggtateee gaetetgetg 240
ctgttcaaaa acggtgaagt ggcggcaacc aaagtgggtg cactgtctaa aggtcagttg 300
aaagagttcc tcgacgctaa cctggccggt tctggttctg gatctccaaa atcggatctg 360
tetggtggtg gtggtggtet ggtteegegt ggateegace egategetgg tgeteaetgg 420
ggtgttctgg ctggtatcgc ttacttctct atggttggta actgggctaa agttctggtt 480
gttctgctgc tgttcgctgg tgttgacgct tag
<210> 36
<211> 117
```



```
<213> Séquence artificielle
<220>
<223> Description de la séquence artificielle: système
     d'expression codant pour la protéine de fusion
     M-DP-TME1
<400> 36
atggaccega tegetggtge teactggggt gttetggetg gtategetta ettetetatg 60
jūtggtaact gggctaaagt tetggttgtt etgetgetgt tegetggtgt tgaeget
<210> 37
<211> 795
:212> ADN
<113> Séquence artificielle
-220>
.223 > Description de la séquence artificielle: système
      d'expression codant pour la protéine de fusion
      GST-DP-TME2
400> 37
atgtccccta tactaggtta ttggaaaatt aagggccttg tgcaacccac tcgacttctt 60
ttggaatatc ttgaagaaaa atatgaagag catttgtatg agcgcgatga aggtgataaa 120
tggcgaaaca aaaagtttga attgggtttg gagtttccca atcttcctta ttatattgat 180
ggtgatgtta aattaacaca gtctatggcc atcatacgtt atatagctga caagcacaac 240
atgttgggtg gttgtccaaa agagcgtgca gagatttcaa tgcttgaagg agcggttttg 300
gatattagat acggtgtttc gagaattgca tatagtaaag actttgaaac tctcaaagtt 360
gattttctta gcaagctacc tgaaatgctg aaaatgttcg aagatcgttt atgtcataaa 420
acatatttaa atggtgatca tgtaacccat cctgacttca tgttgtatga cgctcttgat 480
gttgttttat acatggaccc aatgtgcctg gatgcgttcc caaaattagt ttgttttaaa 540
aaacgtattg aagctatccc acaaattgat aagtacttga aatccagcaa gtatatagca 600
tggcctttgc agggctggca agccacgttt ggtggtggcg accatcctcc aaaatcggat 660
etgtetggtg gtggtggtgg tetggtteeg egtggateeg acceggaata egttgttetg 720
ctgttcctgc tgctggctga cgctcgtgtt tgctcttgcc tgtggatgat gctgctgatc 780
tctcaggctg aagct
<210> 38
<211> 486
<212> ADN
<213> Séquence artificielle
<220>
<223> Description de la séquence artificielle: système
      d'expression codant pour la protéine de fusion
      TrX-DP-TME2
atgagcgata aaattattca cctgactgac gacagttttg acacggatgt actcaaagcg 60
gacggggega teetegtega tttetgggea gagtggtgeg gteegtgeaa aatgategee 120
ccgattctgg atgaaatcgc tgacgaatat cagggcaaac tgaccgttgc aaaactgaac 180
ategateaaa accetggeac tgegeegaaa tatggeatee gtggtateee gactetgetg 240
 ctgttcaaaa acggtgaagt ggcggcaacc aaagtgggtg cactgtctaa aggtcagttg 300
 ggtggtggtc tggttccgcg tggatccgac ccggaatacg ttgttctgct gttcctgctg 420
 ctggctgacg ctcgtgtttg ctcttgcctg tggatgatgc tgctgatctc tcaggctgaa 480
 gcttag
```



```
<210> 39
<211> 99
<212> ADN
<213> Séquence artificielle
<220>
<223> Description de la séquence artificielle: système
     d'expression codant pour la protéine de fusion
     M-DP-TME2
<400> 39
atggacccgg aatacgttgt totgctgttc ctgctgctgg ctgacgctcg tgtttgctct 60
tgcctgtgga tgatgctgct gatctctcag gctgaagct
<210> 40
<211> 5082
<212> ADN
<213> Séquence artificielle
<220>
<223> Description de la séquence artificielle: vecteur
      d'expression pGEXKT-dp-Pt(TME1)
acgttatega etgeacggtg caccaatget tetggegtea ggeagecate ggaagetgtg 60
gtatggctgt gcaggtcgta aatcactgca taattcgtgt cgctcaaggc gcactcccgt 120
totggataat gttttttgcg ccgacatcat aacggttctg gcaaatattc tgaaatgagc 180
tgttgacaat taatcatcgg ctcgtataat gtgtggaatt gtgagcggat aacaatttca 240
cacaggaaac agtattcatg teccetatae taggttattg gaaaattaag ggeettgtge 300
aacccactcg acttettttg gaatatettg aagaaaaata tgaagagcat ttgtatgage 360
gcgatgaagg tgataaatgg cgaaacaaaa agtttgaatt gggtttggag tttcccaatc 420
ttccttatta tattgatggt gatgttaaat taacacagtc tatggccatc atacgttata 480
tagctgacaa gcacaacatg ttgggtggtt gtccaaaaga gcgtgcagag atttcaatgc 540
ttgaaggagc ggttttggat attagatacg gtgtttcgag aattgcatat agtaaagact 600
ttgaaactct caaagttgat tttcttagca agctacctga aatgctgaaa atgttcgaag 660
atogtttatg toataaaaca tatttaaatg gtgatcatgt aacccatcct gacttcatgt 720
tgtatgacgc tcttgatgtt gttttataca tggacccaat gtgcctggat gcgttcccaa 780
aattagtttg ttttaaaaaa cgtattgaag ctatcccaca aattgataag tacttgaaat 840
ccagcaagta tatagcatgg cctttgcagg gctggcaagc cacgtttggt ggtggcgacc 900
atcctccaaa atcggatctg tctggtggtg gtggtggtct ggttccgcgt ggatccgacc 960
cgatcgctgg tgctcactgg ggtgttctgg ctggtatcgc ttacttctct atggttqqta 1020
actgggctaa agttctggtt gttctgctgc tgttcgctgg tgttgacgct taggaattca 1080
tegtgaetga etgaegatet geetegegeg ttteggtgat gaeggtgaaa acetetgaea 1140
catgcagete eeggagaegg teacagettg tetgtaageg gatgeeggga geagaeaage 1200
ccgtcagggc gcgtcagcgg gtgttggcgg gtgtcggggc gcagccatga cccagtcacg 1260
tagcgatagc ggagtgtata attcttgaag acgaaagggc ctcgtgatac gcctattttt 1320
ataggttaat gtcatgataa taatggtttc ttagacgtca ggtggcactt ttcggggaaa 1380
tgtgcgcgga acccctattt gtttattttt ctaaatacat tcaaatatgt atccgctcat 1440
gagacaataa ccctgataaa tgcttcaata atattgaaaa aggaagagta tgagtattca 1500
acatttccgt gtcgccctta ttcccttttt tgcggcattt tgccttcctg tttttgctca 1560
cccagaaacg ctggtgaaag taaaagatgc tgaagatcag ttgggtgcac gagtgggtta 1620
categaactg gateteaaca geggtaagat cettgagagt tttegeeceg aagaacgttt 1680
tecaatgatg ageaetttta aagttetget atgtggegeg gtattateee gtgttgaege 1740
cgggcaagag caactcggtc gccgcataca ctattctcag aatgacttgg ttgagtactc 1800
accagtcaca gaaaagcatc ttacggatgg catgacagta agagaattat gcagtgctgc 1860
cataaccatg agtgataaca ctgcggccaa cttacttctg acaacgatcg gaggaccgaa 1920
ggagctaacc gcttttttgc acaacatggg ggatcatgta actcgccttg atcgttggga 1980
accggagetg aatgaageca taccaaacga cgagegtgae accaegatge etgeageaat 2040
ggcaacaacg ttgcgcaaac tattaactgg cgaactactt actctagett cccggcaaca 2100
attaatagac tggatggagg cggataaagt tgcaggacca cttctgcgct cggcccttcc 2160
```



```
ggctggctgg tttattgctg ataaatctgg agccggtgag cgtgggtctc gcggtatcat 2220
tgcagcactg gggccagatg gtaagccctc ccgtatcgta gttatctaca cgacggggag 2280
tcaggcaact atggatgaac gaaatagaca gatcgctgag ataggtgcct cactgattaa 2340
gcattggtaa ctgtcagacc aagtttactc atatatactt tagattgatt taaaacttca 2400
tttttaattt aaaaggatct aggtgaagat cctttttgat aatctcatga ccaaaatccc 2460
ttaacgtgag ttttcgttcc actgagcgtc agaccccgta gaaaagatca aaggatcttc 2520
ttgagateet tttttetge gegtaatetg etgettgeaa acaaaaaaac caccqctace 2580
ageggtggtt tgtttgeegg atcaagaget accaactett ttteegaagg taaetggett 2640
cagcagagcg cagataccaa atactgtcct tctagtgtag ccgtagttag gccaccactt 2700
caagaactet gtagcacege etacatacet egetetgeta atcetgttae eagtggetge 2760
tgccagtggc gataagtcgt gtcttaccgg gttggactca agacgatagt taccggataa 2820
ggcgcagcgg tcgggctgaa cggggggttc gtgcacacag cccagcttgg agcgaacgac 2880
ctacaccgaa ctgagatacc tacagcgtga gctatgagaa agcgccacgc ttcccgaagg 2940
gagaaaggeg gacaggtate eggtaagegg cagggtegga acaggagage geacgaggga 3000
gcttccaggg ggaaacgcct ggtatettta tagtcctgtc gggtttcgcc acctctgact 3060
tgagcgtcga tttttgtgat gctcgtcagg ggggcggagc ctatggaaaa acgccagcaa 3120
egeggeettt ttaeggttee tggeettttg etggeetttt geteacatgt tettteetge 3180
gttatcccct gattctgtgg ataaccgtat taccgccttt gagtgagctg ataccgctcg 3240
ccgcagccga acgaccgagc gcagcgagtc agtgagcgag gaagcggaag agcgcctgat 3300
geggtatttt ctccttacgc atctgtgcgg tatttcacac cgcataaatt ccgacaccat 3360
cgaatggtgc aaaacetttc gcggtatggc atgatagcgc ccggaagaga gtcaattcag 3420
ggtggtgaat gtgaaaccag taacgttata cgatgtcgca gagtatgccg gtgtctctta 3480
teagacegtt teeegegtgg tgaaceagge cageeaegtt tetgegaaaa egegggaaaa 3540
agtggaagcg gcgatggcgg agctgaatta cattcccaac cgcgtggcac aacaactggc 3600
gggcaaacag tegttgetga ttggcgttge cacetecagt etggceetge acgegeegte 3660
gcaaattgtc gcggcgatta aatctcgcgc cgatcaactg ggtgccagcg tggtggtgtc 3720
gatggtagaa cgaagcggcg tcgaagcctg taaagcggcg gtgcacaatc ttctcgcgca 3780
acgegteagt gggetgatea ttaactatee getggatgae caggatgeca ttgetgtga 3840
agetgeetge actaatgtte eggegttatt tettgatgte tetgaccaga cacceateaa 3900
cagtattatt ttctcccatg aagacggtac gcgactgggc gtggagcatc tggtcgcatt 3960
gggtcaccag caaatcgcgc tgttagcggg cccattaagt tctgtctcgg cgcgtctgcg 4020
totggctggc tggcataaat atotcactcg caatcaaatt cagccgatag cggaacggga 4080
aggcgactgg agtgccatgt ccggttttca acaaaccatg caaatgctga atgagggcat 4140
cgttcccact gcgatgctgg ttgccaacga tcagatggcg ctgggcgcaa tgcgcgccat 4200
taccgagtcc gggctgcgcg ttggtgcgga tatctcggta gtgggatacg acgataccga 4260
agacagetea tgttatatee egeegttaae eaceateaaa eaggatttte geetgetggg 4320
gcaaaccagc gtggaccgct tgctgcaact ctctcagggc caggcggtga agggcaatca 4380
gctgttgccc gtctcactgg tgaaaagaaa aaccaccctg gcgcccaata cgcaaaccgc 4440
ctctccccgc gcgttggccg attcattaat gcagctggca cgacaggttt cccgactgga 4500
aagegggeag tgagegeaac geaattaatg tgagttaget cacteattag geaceceagg 4560
ctttacactt tatgetteeg getegtatgt tgtgtggaat tgtgagegga taacaattte 4620
acacaggaaa cagctatgac catgattacg gattcactgg ccgtcgtttt acaacgtcgt 4680
gactgggaaa accetggegt tacceaett aategeettg cageacatee ceetttegee 4740
agetggegta atagegaaga ggecegeace gategeeett cecaacagtt gegeageetg 4800
aatggcgaat ggcgctttgc ctggtttccg gcaccagaag cggtgccgga aagctggctg 4860
gagtgcgatc ttcctgaggc cgatactgtc gtcgtcccct caaactggca gatgcacggt 4920
tacgatgcgc ccatctacac caacgtaacc tatcccatta cggtcaatcc gccgtttgtt 4980
cccacggaga atccgacggg ttgttactcg ctcacattta atgttgatga aagctggcta 5040
caggaaggcc agacgcgaat tatttttgat ggcgttggaa tt
                                                                   5082
```

```
<210> 41
<211> 5064
<212> ADN
<213> Séquence artificielle
<220>
<223> Description de la séquence artificielle: vecteur d'expression pGEXKT-dp-Pt(TME2)
```

<400> 41



acgttatega etgeaeggtg caccaatget tetggegtea ggeageeate ggaagetgtg 60 gtatggctgt gcaggtcgta aatcactgca taattcgtgt cgctcaaggc gcactcccgt 120 totggataat gttttttgcg ccgacatcat aacggttctg gcaaatattc tgaaatgagc 180 tgttgacaat taatcatcgg ctcgtataat gtgtggaatt gtgagcggat aacaatttca 240 cacaggaaac agtattcatg tcccctatac taggttattg gaaaattaag ggccttgtgc 300 aacccactcg acttcttttg gaatatcttg aagaaaaata tgaagagcat ttgtatgagc 360 gcgatgaagg tgataaatgg cgaaacaaaa agtttgaatt gggtttggag tttcccaatc 420 ttccttatta tattgatggt gatgttaaat taacacagtc tatggccatc atacgttata 480 tagetgacaa geacaacatg ttgggtggtt gtecaaaaga gegtgeagag attteaatge 540 ttgaaggagc ggttttggat attagatacg gtgtttcgag aattgcatat agtaaagact 600 ttgaaactct caaagttgat tttcttagca agctacctga aatgctgaaa atgttcgaag 660 atcgtttatg tcataaaaca tatttaaatg gtgatcatgt aacccatcct gacttcatgt 720 tgtatgacgc tcttgatgtt gttttataca tggacccaat gtgcctggat gcgttcccaa 780 aattagtttg ttttaaaaaa cgtattgaag ctatcccaca aattgataag tacttgaaat 840 ccagcaagta tatagcatgg cctttgcagg gctggcaagc cacgtttggt ggtggcgacc 900 atcctccaaa atcggatctg tctggtggtg gtggtggtct ggttccgcgt ggatccgacc 960 eggaataegt tgttetgetg tteetgetge tggetgaege tegtgtttge tettgeetgt 1020 ggatgatgct gctgatctct caggctgaag cttaggaatt catcgtgact gactgacgat 1080 ctgcctcgcg cgtttcggtg atgacggtga aaacctctga cacatgcagc tcccggagac 1140 ggtcacaget tgtctgtaag cggatgccgg gagcagacaa gcccgtcagg gcgcgtcagc 1200 gggtgttggc gggtgtcggg gcgcagccat gacccagtca cgtagcgata gcggagtgta 1260 taattettga agacgaaagg geetegtgat aegeetattt ttataggtta atgteatgat 1320 aataatggtt tettagaegt eaggtggeae ttttegggga aatgtgegeg gaaceeetat 1380 ttgtttattt ttctaaatac attcaaatat gtatccgctc atgagacaat aaccctgata 1440 aatgottcaa taatattgaa aaaggaagag tatgagtatt caacatttcc gtgtcgccct 1500 tattecettt tttgeggeat tttgeettee tgtttttget eacceagaaa egetggtgaa 1560 agtaaaagat gctgaagatc agttgggtgc acgagtgggt tacatcgaac tggatctcaa 1620 cageggtaag atcettgaga gttttegece egaagaaegt tttecaatga tgageaettt 1680 taaagttetg etatgtggeg eggtattate eegtgttgae geegggeaag ageaactegg 1740 togoogoata cactattoto agaatgactt ggttgagtac toaccagtoa cagaaaagca 1800 tettaeggat ggeatgaeag taagagaatt atgeagtget geeataacea tgagtgataa 1860 cactgcggcc aacttacttc tgacaacgat cggaggaccg aaggagctaa ccgcttttt 1920 gcacaacatg ggggatcatg taactcgcct tgatcgttgg gaaccggagc tgaatgaagc 1980 cataccaaac gacgagcgtg acaccacgat gcctgcagca atggcaacaa cgttgcgcaa 2040 actattaact ggcgaactac ttactctagc ttcccggcaa caattaatag actggatgga 2100 ggcggataaa gttgcaggac cacttctgcg ctcggccctt ccggctggct ggtttattgc 2160 tgataaatct ggagccggtg agcgtgggtc tcgcggtatc attgcagcac tggggccaga 2220 tggtaagccc tcccgtatcg tagttatcta cacgacgggg agtcaggcaa ctatggatga 2280 acgaaataga cagatcgctg agataggtgc ctcactgatt aagcattggt aactgtcaga 2340 ccaagtttac tcatatatac tttagattga tttaaaactt catttttaat ttaaaaggat 2400 ctaggtgaag atcettttg ataateteat gaccaaaate cettaacgtg agttttegtt 2460 ccactgagcg tcagaccccg tagaaaagat caaaggatct tcttgagatc ctttttttct 2520 gcgcgtaatc tgctgcttgc aaacaaaaaa accaccgcta ccagcggtgg tttgtttgcc 2580 ggatcaagag ctaccaactc tttttccgaa ggtaactggc ttcagcagag cgcagatacc 2640 aaatactgtc cttctagtgt agccgtagtt aggccaccac ttcaagaact ctgtagcacc 2700 gectacatac etegetetge taateetgtt accagtgget getgecagtg gegataagte 2760 gtgtcttacc gggttggact caagacgata gttaccggat aaggcgcagc ggtcgggctg 2820 aacggggggt tcgtgcacac agcccagctt ggagcgaacg acctacaccg aactgagata 2880 cctacagcgt gagctatgag aaagcgccac gcttcccgaa gggagaaagg cggacaggta 2940 teeggtaage ggeagggteg gaacaggaga gegeacgagg gagetteeag ggggaaaege 3000 ctggtatctt tatagtcctg tcgggtttcg ccacctctga cttgagcgtc gatttttgtg 3060 atgetegtea ggggggegga geetatggaa aaaegeeage aaegeggeet ttttaeggtt 3120 cetggeettt tgetggeett ttgeteacat gttettteet gegitateee etgattetgt 3180 ggataaccgt attaccgcct ttgagtgagc tgataccgct cgccgcagcc gaacgaccga 3240 gcgcagcgag tcagtgagcg aggaagcgga agagcgcctg atgcggtatt ttctccttac 3300 gcatctgtgc ggtatttcac accgcataaa ttccgacacc atcgaatggt gcaaaacctt 3360 tegeggtatg geatgatage geceggaaga gagteaatte agggtggtga atgtgaaace 3420 agtaacgtta tacgatgtcg cagagtatgc cggtgtctct tatcagaccg tttcccgcgt 3480 ggtgaaccag gccagccacg tttctgcgaa aacgcgggaa aaagtggaag cggcgatggc 3540 ggagctgaat tacattccca accgcgtggc acaacaactg gcgggcaaac agtcgttgct 3600 gattggcgtt gccacetcca gtctggccct gcacgcgccg tcgcaaattg tcgcggcgat 3660



```
taaatetege geegateaae tgggtgeeag egtggtggtg tegatggtag aaegaagegg 3720
cgtcgaagcc tgtaaagcgg cggtgcacaa tcttctcgcg caacgcgtca gtgggctgat 3780
cattaactat ccgctggatg accaggatgc cattgctgtg gaagctgcct gcactaatgt 3840
tccggcgtta tttcttgatg tctctgacca gacacccatc aacagtatta ttttctccca 3900
tgaagacggt acgcgactgg gcgtggagca tctggtcgca ttgggtcacc agcaaatcgc 3960
gctgttagcg ggcccattaa gttctgtctc ggcgcgtctg cgtctggctg gctggcataa 4020
atateteact egeaateaaa tteageegat ageggaaegg gaaggegaet qqaqtqccat 4080
gtccggtttt caacaaacca tgcaaatgct gaatgaggc atcgttccca ctgcgatgct 4140
ggttgccaac gatcagatgg cgctgggcgc aatgcgcgcc attaccgagt ccgggctgcg 4200
cgttggtgcg gatatctcgg tagtgggata cgacgatacc gaagacagct catgttatat 4260
cccgccgtta accaccatca aacaggattt tcgcctgctg gggcaaacca gcgtggaccg 4320
ettgetgeaa eteteteagg geeaggeggt gaagggeaat cagetgttge cegteteaet 4380
ggtgaaaaga aaaaccaccc tggcgcccaa tacgcaaacc gcctctcccc gcgcgttggc 4440
cgattcatta atgcagctgg cacgacaggt ttcccgactg gaaagcgggc agtgagcgca 4500
acgcaattaa tgtgagttag ctcactcatt aggcacccca ggctttacac tttatgcttc 4560
eggetegtat gttgtgtgga attgtgageg gataacaatt teacacagga aacagetatg 4620
accatgatta eggattoact ggeegtegtt ttacaaegte gtgaetggga aaaccetgge 4680
gttacccaac ttaatcgcct tgcagcacat ccccctttcg ccagctggcg taatagcgaa 4740
gaggcccgca ccgatcgccc ttcccaacag ttgcgcagcc tgaatggcga atggcgcttt 4800
geetggttte eggeaceaga ageggtgeeg gaaagetgge tggagtgega tetteetgag 4860
 geogatactg tegtegtece eteaaactgg cagatgeacg_gttacgatge geocatetae 4920
 accaacgtaa cetateceat taeggteaat eegeegtttg tteecaegga gaateegaeg 4980
 ggttgttact cgctcacatt taatgttgat gaaagctggc tacaggaagg ccagacgcga 5040
 attatttttg atggcgttgg aatt
 <210> 42
 <211> 5918
 <212> ADN
 <213> Séquence artificielle
 <220>
```

```
<223> Description de la séquence artificielle: vecteur
     d'expression pET32a-dp-Pt(TME1)
```

<400> 42 atcoggatat agttoctcot ttcagcaaaa aaccoctcaa gaccogttta gaggcoccaa 60 ggggttatgc tagttattgc tcagcggtgg cagcagccaa ctcagcttcc tttcgggctt 120 tgttagcagc cggatctcag tggtggtggt ggtggtgctc gagtgcggcc gcaagcttgt 180 cgacggagct cgaattccta agcgtcaaca ccagcgaaca gcagcagaac aaccagaact 240 ttagcccagt taccaaccat agagaagtaa gcgataccag ccagaacacc ccagtgagca 300 ccagcgatcg ggtcggatcc acgcggaacc agaccaccac caccaccaga cagatccgat 360 tttggagatc cagaaccaga accggccagg ttagcgtcga ggaactcttt caactgacct 420 ttagacagtg cacccacttt ggttgccgcc acttcaccgt ttttgaacag cagcagagtc 480 gggataccac ggatgccata tttcggcgca gtgccagggt tttgatcgat gttcagtttt 540 gcaacggtca gtttgccctg atattcgtca gcgatttcat ccagaatcgg ggcgatcatt 600 ttgcacggac cgcaccactc tgcccagaaa tcgacgagga tcgccccgtc cgctttgagt 660 acatecgtgt caaaactgte gteagteagg tgaataattt tategeteat atgtatatet 720 ccttcttaaa gttaaacaaa attatttcta gaggggaatt gttatccgct cacaattccc 780 ctatagtgag tcgtattaat ttcgcgggat cgagatcgat ctcgatcctc tacgccggac 840 gcatcgtggc cggcatcacc ggcgccacag gtgcggttgc tggcgcctat atcgccgaca 900 teacegatgg ggaagategg getegeeact tegggeteat gagegettgt tteggegtgg 960 gtatggtggc aggccccgtg gccgggggac tgttgggcgc catctccttg catgcaccat 1020 teettgegge ggeggtgete aacggeetea acetactact gggetgette etaatgeagg 1080 agtegeataa gggagagegt egagateeeg gacaceateg aatggegeaa aacetttege 1140 ggtatggcat gatagcgccc ggaagagagt caattcaggg tggtgaatgt gaaaccagta 1200 acgttatacg atgtcgcaga gtatgccggt gtctcttatc agaccgtttc ccgcgtggtg 1260 aaccaggcca gccacgtttc tgcgaaaacg cgggaaaaag tggaagcggc gatggcggag 1320 ctgaattaca ttcccaaccg cgtggcacaa caactggcgg gcaaacagtc gttgctgatt 1380 ggcgttgcca cctccagtct ggccctgcac gcgccgtcgc aaattgtcgc ggcgattaaa 1440 tetegegeeg ateaactggg tgecagegtg gtggtgtega tggtagaacg aageggegte 1500



gaagcctgta aagcggcggt gcacaatctt ctcgcgcaac gcgtcagtgg gctgatcatt 1560 aactatccgc tggatgacca ggatgccatt gctgtggaag ctgcctgcac taatgttccg 1620 gcgttatttc ttgatgtctc tgaccagaca cccatcaaca gtattatttt ctcccatgaa 1680 gacggtacgc gactgggcgt ggagcatctg gtcgcattgg gtcaccagca aatcgcgctg 1740 ttagcgggcc cattaagttc tgtctcggcg cgtctgcgtc tggctggctg gcataaatat 1800 ctcactcgca atcaaattca gccgatagcg gaacgggaag gcgactggag tgccatgtcc 1860 ggttttcaac aaaccatgca aatgctgaat gagggcatcg ttcccactgc gatgctggtt 1920 gccaacgatc agatggcgct gggcgcaatg cgcgccatta ccgagtccgg gctgcgcgtt 1980 ggtgcggaca tctcggtagt gggatacgac gataccgaag acagctcatg ttatatcccg 2040 ccgttaacca ccatcaaaca ggattttcgc ctgctggggc aaaccagcgt ggaccgcttg 2100 ctgcaactct ctcagggcca ggcggtgaag ggcaatcagc tgttgcccgt ctcactggtg 2160 aaaagaaaaa ccacctggc gcccaatacg caaaccgcct ctccccgcgc gttggccgat 2220 tcattaatgc agctggcacg acaggtttcc cgactggaaa gcggcagtg agcgcaacgc 2280 aattaatgta agttagctca ctcattaggc accgggatct cgaccgatgc ccttgagagc 2340 cttcaaccca gtcagctcct tccggtgggc gcggggcatg actatcgtcg ccgcacttat 2400 gactgtcttc tttatcatgc aactcgtagg acaggtgccg gcagcgctct gggtcatttt 2460 cggcgaggac cgctttcgct ggagcgcgac gatgatcggc ctgtcgcttg cggtattcgg 2520 aatottgcac geeetegete aageettegt caetggteee geeaccaaac gttfeggega 2580 gaagcaggcc attategeeg geatggegge eccaegggtg egeatgateg tgeteetgte 2640 gttgaggacc cggctaggct ggcggggttg ccttactggt tagcagaatg aatcaccgat 2700 acgcgagcga acgtgaagcg actgctgctg caaaacgtct gcgacctgag caacaacatg 2760 aatggtette ggttteegtg tttegtaaag tetggaaaeg eggaagteag egeeetgeae 2820 cattatgtte eggatetgea tegeaggatg etgetggeta ceetgtggaa cacetacate 2880 tgtattaacg aagcgctggc attgaccctg agtgattttt ctctggtccc gccgcatcca 2940 taccgccagt tgtttaccct cacaacgttc cagtaaccgg gcatgttcat catcagtaac 3000 ccgtatcgtg agcatcctct ctcgtttcat cggtatcatt acccccatga acagaaatcc 3060 cccttacacg gaggcatcag tgaccaaaca ggaaaaaacc gcccttaaca tggcccgctt 3120 tatcagaagc cagacattaa cgcttctgga gaaactcaac gagctggacg cggatgaaca 3180 ggcagacatc tgtgaatcgc ttcacgacca cgctgatgag ctttaccgca gctgcctcgc 3240 gcgtttcggt gatgacggtg aaaacctctg acacatgcag ctcccggaga cggtcacagc 3300 ttgtctgtaa gcggatgccg ggagcagaca agcccgtcag ggcgcgtcag cgggtgttgg 3360 cgggtgtcgg ggcgcagcca tgacccagtc acgtagcgat agcggagtgt atactggctt 3420 aactatgcgg catcagagca gattgtactg agagtgcacc atatatgcgg tgtgaaatac 3480 cgcacagatg cgtaaggaga aaataccgca tcaggcgctc ttccgcttcc tcgctcactg 3540 actogotgog ctoggtogtt oggotgoggo gagoggtato agotoactoa aaggoggtaa 3600 tacggttatc cacagaatca ggggataacg caggaaagaa catgtgagca aaaggccagc 3660 aaaaggccag gaaccgtaaa aaggccgcgt tgctggcgtt tttccatagg ctccgcccc 3720 ctgacgagca tcacaaaaat cgacgctcaa gtcagaggtg gcgaaacccg acaggactat 3780 aaagatacca ggcgtttccc cctggaagct ccctcgtgcg ctctcctgtt ccgaccctgc 3840 cgcttaccgg atacctgtcc gcctttctcc cttcgggaag cgtggcgctt tctcatagct 3900 cacgctgtag gtatctcagt tcggtgtagg tcgttcgctc caagctgggc tgtgtgcacg 3960 aaccccccgt tcagcccgac cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc 4020 cggtaagaca cgacttatcg ccactggcag cagccactgg taacaggatt agcagagcga 4080 ggtatgtagg cggtgctaca gagttcttga agtggtggcc taactacggc tacactagaa 4140 ggacagtatt tggtatctgc gctctgctga agccagttac cttcggaaaa agagttggta 4200 gctcttgatc cggcaaacaa accaccgctg gtagcggtgg tttttttgtt tgcaagcagc 4260 agattacgcg cagaaaaaaa ggatctcaag aagatccttt gatcttttct acggggtctg 4320 acgctcagtg gaacgaaaac tcacgttaag ggattttggt catgagatta tcaaaaagga 4380 tetteaceta gateetttta aattaaaaat gaagttttaa atcaatetaa agtatatatg 4440 agtaaacttg gtctgacagt taccaatgct taatcagtga ggcacctatc tcagcgatct 4500 gtctatttcg ttcatccata gttgcctgac tccccgtcgt gtagataact acgatacggg 4560 agggettace atetggeece agtgetgeaa tgatacegeg agacecaege teaceggete 4620 cagatttatc agcaataaac cagccagccg gaagggccga gcgcagaagt ggtcctgcaa 4680 ctttatccgc ctccatccag tctattaatt gttgccggga agctagagta agtagttcgc 4740 cagttaatag tttgcgcaac gttgttgcca ttgctgcagg catcgtggtg tcacgctcgt 4800 cgtttggtat ggcttcattc agctccggtt cccaacgatc aaggcgagtt acatgatccc 4860 ccatgttgtg caaaaaagcg gttagctcct tcggtcctcc gatcgttgtc agaagtaagt 4920 tggccgcagt gttatcactc atggttatgg cagcactgca taattctctt actgtcatgc 4980 catccgtaag atgcttttct gtgactggtg agtactcaac caagtcattc tgagaatagt 5040 gtatgeggeg accgagttge tettgeeegg egteaatacg ggataatace gegeeacata 5100 gcagaacttt aaaagtgctc atcattggaa aacgttcttc ggggcgaaaa ctctcaagga 5160 ·



```
tettaceget gttgagatee agttegatgt aacecacteg tgcacecaac tgatetteag 5220 catettttac ttteaccage gtttetgggt gagcaaaaac aggaaggcaa aatgeegcaa 5280 aaaagggaat aagggegaca eggaaatgtt gaatacteat actetteett ttteaatatt 5340 attgaaggat ttategetea tgageggata catattegaa tgtatttaga 5400 aaaataaaca aataggggtt eegegacat tteecegaaa agtgeeacet gaaattgtaa 5460 acgttaatat tttgttaaaa tteegegtaa attettgta aatecagetea tttttaace 5520 aataggeega aateggeaaa atecettata aateaaaaga atagaeegag atagggttga 5580 gtgttgttee agtttggaac aagagteeac tattaaagaa egggaeege aaeggegeege eactaegtga accateacee taateaagtt 5700 gagettggate gaggtgeegt aaageacetaa ateggaaeee taaagggage cecegattta 5760 gagettgaeg gegaaegtgg egagaaagga agggaagaaa gegaaaggag 5820 egggeegtaa gegetggea agtgtagegg teaegetgee egtaaeee 5918
```

```
-#10> 43
```

<211> 5891

312> ADN

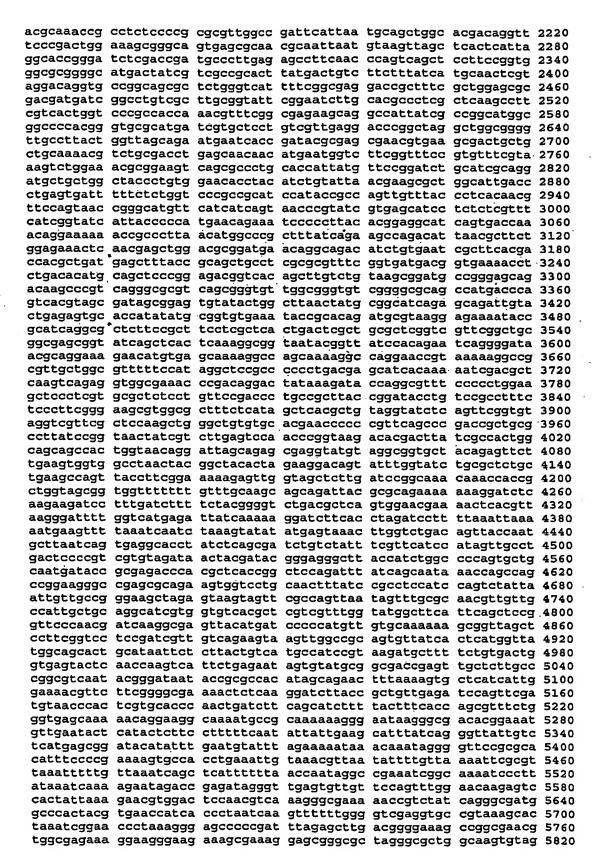
413> Séquence artificielle

<220>

<223> Description de la séquence artificielle: vecteur d'expression pET32a-dp-Pt(TME2)

<400> 43 atccggatat agttcctcct ttcagcaaaa aacccctcaa gacccgttta gaggccccaa 60 ggggttatgc tagttattgc tcagcggtgg cagcagccaa ctcagcttcc tttcgggctt 120 tgttagcagc cggatctcag tggtggtggt ggtggtgctc gagtgcggcc gcaagcttgt 180 cgacggaget cgaatteeta agetteagee tgagagatea geageateat ceacaggeaa 240 gagcaaacac gagcgtcagc cagcagcagg aacagcagaa caacgtattc cgggtcggat 300 ccacgeggaa ccagaccacc accaccacca gacagatcag atccagaacc agaaccggcc 360 aggttagegt egaggaacte ttteaactga cetttagaca gtgcacceae tttggttgee 420 gccacttcac cgtttttgaa cagcagcaga gtcgggatac cacggatgcc atatttcggc 480 gcagtgccag ggttttgatc gatgttcagt tttgcaacgg tcagtttgcc ctgatattcg 540 tcagcgattt catccagaat cggggcgatc attttgcacg gaccgcacca ctctgcccag 600 aaategacga ggategeece gteegetttg agtacateeg tgteaaaact gtegteagte 660 aggtgaataa ttttatcgct catatgtata tctccttctt aaagttaaac aaaattattt 720 ctagagggga attgttatcc gctcacaatt cccctatagt gagtcgtatt aatttcgcgg 780 gategagate gatetegate etetaegeeg gaegeategt ggeeggeate aceggegeea 840 caggtgcggt tgctggcgcc tatatcgccg acatcaccga tggggaagat cgggctcgcc 900 actteggget catgageget tgttteggeg tgggtatggt ggcaggeece gtggeegggg 960 gactgttggg cgccatctcc ttgcatgcac cattccttgc ggcggcggtg ctcaacggcc 1020 tcaacctact actgggctgc ttcctaatgc aggagtcgca taagggagag cgtcgagatc 1080 ccggacacca tcgaatggcg caaaaccttt cgcggtatgg catgatagcg cccggaagag 1140 agtcaattca gggtggtgaa tgtgaaacca gtaacgttat acgatgtcgc agagtatgcc 1200 ggtgtctett atcagaccgt ttcccgcgtg gtgaaccagg ccagccacgt ttctgcgaaa 1260 acgegggaaa aagtggaage ggegatggeg gagetgaatt acatteecaa eegegtggea 1320. caacaactgg cgggcaaaca gtcgttgctg attggcgttg ccacctccag tctggccctg 1380 cacgegeegt egeaaattgt egeggegatt aaatetegeg eegateaaet gggtgeeage 1440 gtggtggtgt cgatggtaga acgaagcggc gtcgaagcct gtaaagcggc ggtgcacaat 1500 ettetegege aacgegteag tgggetgate attaactate egetggatga ccaggatgee 1560 attgctgtgg aagctgcctg cactaatgtt ccggcgttat ttcttgatgt ctctgaccag 1620 acacccatca acagtattat tttctcccat gaagacggta cgcgactggg cgtggagcat 1680 ctggtcgcat tgggtcacca gcaaatcgcg ctgttagcgg gcccattaag ttctgtctcg 1740 gegegtetge gtetggetgg etggeataaa tateteacte geaateaaat teageegata 1800 geggaacggg aaggegactg gagtgecatg teeggtttte aacaaaccat geaaatgetg 1860 aatgagggca tegtteecae tgegatgetg gttgeeaacg atcagatgge getgggegca 1920 atgegegeca ttacegagte egggetgege gttggtgegg acateteggt agtgggatae 1980 gacgataccg aagacagete atgttatate eegeegttaa eeaccateaa acaggatttt 2040 cgcctgctgg ggcaaaccag cgtggaccgc ttgctgcaac tctctcaggg ccaggcggtg 2100 aagggcaate agetgttgee egteteactg gtgaaaagaa aaaccaccet ggegeecaat 2160









cggtcacgct gcgcgtaacc accacacccg ccgcgcttaa tgcgccgcta cagggcgcgt 5880 cccattcgcc a 5891

<210> 44

<211> 2617

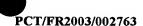
<212> ADN

<213> Séquence artificielle

· <220>

<223> Description de la séquence artificielle: vecteur
 d'expression pT7-7-dp-Pt(TME1)

<400> 44 aatteteatg titgacaget tateategat gataagettg ggetgeaggt egactetaga 60 ggatccccgg gcgcgaattc ctaagcgtca acaccagcga acagcagcag aacaaccaga 120 actttagccc agttaccaac catagagaag taagcgatac cagccagaac accccagtga 180 gcaccagcga tcgggtccat atgtatatct ccttcttaaa gttaaacaaa attatttcta 240 gagggaaacc gttgtggtct ccctatagtg agtcgtatta atttcgaagt ctatcagaag 300 ttegaatege tgggeetege gegttteggt gatgaeggtg aaaacetetg acacatgeag 360 ctcccggaga cggtcacagc ttgtctgtaa gcggatgccg ggagcagaca agcccgtcag 420 ggcgcgtcag cgggtgttgg cgggtgtcgg ggcgcagcca tgacccagtc acgtagcgat 480 agcggagtgt atatactggc ttaactatgc ggcatcagag cagattgtac tgagagtgca 540 ccataggaag atcttccgga agatcttcct atgcggtgtg aaataccgca cagatgcgta 600 aggagaaaat accgcatcag gcgctcttcc gcttcctcgc tcactgactc gctgcgctcg 660 gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg cggtaatacg gttatccaca 720 gaatcagggg ataacgcagg aaagaacatg tgagcaaaaa gccagcaaaa ggccaggaac 780 cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc gccccctga cgagcatcac 840 aaaaatcgac gctcaagtca gaggtggcga aacccgacag gactataaag ataccaggcg 900 tttccccctg gaageteect egtgegetet cetgtteega eeetgeeget taceggatae 960 etgteegeet tteteeette gggaagegtg gegetttete aatgeteacg etgtaggtat 1020 cteagttegg tgtaggtegt tegeteeaag etgggetgtg tgeacgaace eccegtteag 1080 cccgaccgct gcgccttatc cggtaactat cgtcttgagt ccaacccggt aagacacgac 1140 ttategecac tggcagcage cactggtaac aggattagca gagcgaggta tgtaggcggt 1200 gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaaggac agtatttggt 1260 atotgogoto tgotgaagoo agttacotto ggaaaaagag ttggtagoto ttgatocggo 1320 aaacaaacca ccgctggtag cggtggtttt tttgtttgca agcagcagat tacgcgcaga 1380 aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc tcagtggaac 1440 gaaaactcac gttaagggat tttggtcatg agattatcaa aaaggatctt cacctagatc 1500 cttttaattc ttgaagacga aagggcctcg tgatacgcct atttttatag gttaatgtca 1560 tgataataat ggtttcttag acgtcaggtg gcacttttcg gggaaatgtg cgcggaaccc 1620 ctatttgttt atttttctaa atacattcaa atatgtatcc gctcatgaga caataaccct 1680 gataaatgct tcaataatat tgaaaaagga agagtatgag tattcaacat ttccgtgtcg 1740 cccttattcc cttttttgcg gcattttgcc ttcctgtttt tgctcaccca gaaacgctgg 1800 tgaaagtaaa agatgctgaa gatcagttgg gtgcacgagt gggttacatc gaactggatc 1860 tcaacagcgg taagatcctt gagagttttc gccccgaaga acgttttcca atgatgagca 1920 cttttaaagt tetgetatgt ggegeggtat tatecegtgt tgaegeeggg caagageaac 1980 teggtegeeg catacactat teteagaatg aettegttga gtaeteacea gteacagaaa 2040 agcatcttac ggatggcatg acagtaagag aattatgcag tgctgccata accatgagtg 2100 ataacactgc ggccaactta cttctgacaa cgatcggagg accgaaggag ctaaccgctt 2160 ttttgcacaa catgggggat catgtaactc gccttgatcg ttgggaaccg gagctgaatg 2220 aagecataee aaaegaegag egtgaeaeea egatgeetgt ageaatggea acaaegttge 2280 gcaaactatt aactggcgaa ctacttactc tagcttcccg gcaacaatta atagactgga 2340 tggaggcgga taaagttgca ggaccacttc tgcgctcggc ccttccggct ggctggttta 2400 ttgctgataa atctggagcc ggtgagcgtg ggtctcgcgg tatcattgca gcactggggc 2460 cagatggtaa gccctcccgt atcgtagtta tctacacgac ggggagtcag gcaactatgg 2520 atgaacgaaa tagacagatc gctgagatag gtgcctcact gattaagcat tggtaactgt 2580 cagaccaagt ttactcatat atactttaga ttgattt



```
<211> 2599
<212> ADN
<213> Séquence artificielle
<220>
<223> Description de la séquence artificielle: vecteur
      d'expression pT7-7-dp-Pt(TME2)
<400> 45
aattotoatg titgacagot tatoatogat gataagottg ggotgcaggt cgactotaga 60
ggatccccgg gcgcgaattc ctaagcttca gcctgagaga tcagcagcat catccacagg 120
caagagcaaa cacgagcgtc agccagcagc aggaacagca gaacaacgta ttccgggtcc 180
atatgtatat ctccttctta aagttaaaca aaattatttc tagagggaaa ccgttgtggt 240
ctccctatag tgagtcgtat taatttcgaa gtctatcaga agttcgaatc gctgggcctc 300
gegegttteg gtgatgaegg tgaaaacete tgacacatge ageteeegga gaeggteaca 360
gettgtetgt aageggatge egggageaga caagecegte aggggegegte agegggtgtt 420
ggcgggtgtc ggggcgcagc catgacccag tcacgtagcg atagcggagt gtatatactg 480
gcttaactat gcggcatcag agcagattgt actgagagtg caccatagga agatcttccg 540
gaagatette etatgeggtg tgaaataceg cacagatgeg taaggagaaa atacegcate 600
aggegetett eegetteete geteaetgae tegetgeget eggtegtteg getgeggega 660
geggtateag etcaetcaaa ggeggtaata eggttateea cagaatcagg ggataacgca 720
ggaaagaaca tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa ggccgcgttg 780
ctggcgtttt tccataggct ccgccccct gacgagcatc acaaaaatcg acgctcaagt 840
cagaggtggc gaaacccgac aggactataa agataccagg cgtttccccc tggaagctcc 900
ctcgtgcgct ctcctgttcc gaccctgccg cttaccggat acctgtccgc ctttctccct 960
tegggaageg tggegettte teaatgetea egetgtaggt ateteagtte ggtgtaggte 1020
gttcgctcca agctgggctg tgtgcacgaa ccccccgttc agcccgaccg ctgcgcctta 1080
teeggtaaet ategtettga gteeaaceeg gtaagacaeg aettategee aetggeagea 1140
gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga gttcttgaag 1200
tggtggccta actacggcta cactagaagg acagtatttg gtatctgcgc tctgctgaag 1260
ccagttacct toggaaaaag agttggtagc tottgatccg gcaaacaaac caccgctggt 1320
ageggtggtt tttttgtttg caageageag attacgegea gaaaaaaagg ateteaagaa 1380
gateetttga tettttetae ggggtetgae geteagtgga acgaaaaete acgttaaggg 1440
attttggtca tgagattatc aaaaaggatc ttcacctaga tccttttaat tcttgaagac 1500
gaaagggcct cgtgatacgc ctatttttat aggttaatgt catgataata atggtttctt 1560
agacgtcagg tggcactttt cggggaaatg tgcgcggaac ccctatttgt ttattttct 1620
aaatacattc aaatatgtat ccgctcatga gacaataacc ctgataaatg cttcaataat 1680
attgaaaaag gaagagtatg agtattcaac atttccgtgt cgcccttatt cccttttttg 1740
cggcattttg ccttcctgtt tttgctcacc cagaaacgct ggtgaaagta aaagatgctg 1800
aagatcagtt gggtgcacga gtgggttaca tcgaactgga tctcaacagc ggtaagatcc 1860
ttgagagttt tcgccccgaa gaacgttttc caatgatgag cacttttaaa gttctgctat 1920
gtggcgcggt attateccgt gttgacgccg ggcaagagca actcggtcgc cgcatacact 1980
attotoagaa tgacttggtt gagtactcac cagtcacaga aaagcatott acggatggca 2040
tgacagtaag agaattatgc agtgctgcca taaccatgag tgataacact gcggccaact 2100
tacttctgac aacgatcgga ggaccgaagg agctaaccgc ttttttgcac aacatggggg 2160
atcatgtaac tcgccttgat cgttgggaac cggagctgaa tgaagccata ccaaacgacg 2220
agcgtgacac cacgatgcct gtagcaatgg caacaacgtt gcgcaaacta ttaactggcg 2280
aactacttac tctagcttcc cggcaacaat taatagactg gatggaggcg gataaagttg 2340
caggaccact tetgegeteg gecetteegg etggetggtt tattgetgat aaatetggag 2400
ccggtgagcg tgggtctcgc ggtatcattg cagcactggg gccagatggt aagccctccc 2460
gtatcgtagt tatctacacg acggggagtc aggcaactat ggatgaacga aatagacaga 2520
togotgagat aggtgootca otgattaago attggtaact gtoagaccaa gtttactcat 2580
atatacttta gattgattt
                                                                  2599
<210> 46
<211> 271
<212> PRT
<213> Séquence artificielle
<220>
```



<223> Description de la séquence artificielle: protéine de fusion GST-DP-TME1

## <400> 46

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro 1 5 10 15

Thr Arg Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu 20 25 30.

Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu 35 40 45

Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys
50 55 60

Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn 65 70 75 80

Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu 85. 90 95

Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser 100 105 110

Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu
115 120 125

Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn 130 135 140

Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp 145 150 155 160

Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu
165 170 175

Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr 180 185 190

Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala 195 200 205

Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Ser Gly Gly 210 215 220

Gly Gly Gly Leu Val Pro Arg Gly Ser Asp Pro Ile Ala Gly Ala His 225 230 240

Trp Gly Val Leu Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp
245 250 255

Ala Lys Val Leu Val Val Leu Leu Leu Phe Ala Gly Val Asp Ala 260 265 270

<sup>&</sup>lt;210> 47

<sup>&</sup>lt;211> 265

<sup>&</sup>lt;212> PRT

<sup>&</sup>lt;213> Séquence artificielle



<220>

<223> Description de la séquence artificielle: protéine de fusion GST-DP-TME2

<400> 47

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro
1 5 10 15

Thr Arg Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu 20 25 30

Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu 35 40 45

Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys
50 60

Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp'Lys His Asn 65 70 75 80

Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu 85 90 95

Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser 100 105 110

Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu 115 120 125

Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn 130 135 140

Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp 145 150 155 160

Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu 165 170 175

Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr 180 185 190

Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala 195 200 205

Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Ser Gly Gly 210 220

Gly Gly Gly Leu Val Pro Arg Gly Ser Asp Pro Glu Tyr Val Val Leu 225 230 235 240

Leu Phe Leu Leu Ala Asp Ala Arg Val Cys Ser Cys Leu Trp Met 245 250 255

Met Leu Leu Ile Ser Gln Ala Glu Ala 260 265

<210> 48

<211> 170

<212> PRT

<213> Séquence artificielle



<220>

<223> Description de la séquence artificielle: protéine de fusion TrX-DP-TME1

<400> 48

Met Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser Phe Asp Thr Asp 1 5 10 15

Tal Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe Trp Ala Glu Trp
20 25 30

Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp 35 40 45

oiu Tyr Gln Gly Lys Leu Thr Val Ala Lys Leu Asn Ile Asp Gln Asn 50 55 60

10 Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu 55 75 80

Leu Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val Gly Ala Leu Ser 85 90 95

Lys Gly Gln Leu Lys Glu Phe Leu Asp Ala Asn Leu Ala Gly Ser Gly
100 105 110

Ser Gly Ser Pro Lys Ser Asp Leu Ser Gly Gly Gly Gly Leu Val 115 120 125

Pro Arg Gly Ser Asp Pro Ile Ala Gly Ala His Trp Gly Val Leu Ala 130 135 140

Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val Leu Val 145 150 155 160

Val Leu Leu Phe Ala Gly Val Asp Ala 165 170

<210> 49

<211> 161

<212> PRT

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: protéine de fusion TrX-DP-TME2

<400> 49

Met Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser Phe Asp Thr Asp 1 5 10 15

Val Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe Trp Ala Glu Trp 20 25 30

Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp
35 40 45

Glu Tyr Gln Gly Lys Leu Thr Val Ala Lys Leu Asn Ile Asp Gln Asn 50 55 60

## WO 2004/027068



Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu 65 70 75 80

Leu Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val Gly Ala Leu Ser 85 90 95

Lys Gly Gln Leu Lys Glu Phe Leu Asp Ala Asn Leu Ala Gly Ser Gly
100 105 110

Ser Gly Ser Asp Leu Ser Gly Gly Gly Gly Leu Val Pro Arg Gly 115 120 125

Ser Asp Pro Glu Tyr Val Val Leu Leu Phe Leu Leu Leu Ala Asp Ala 130 135 140

Arg Val Cys Ser Cys Leu Trp Met Met Leu Leu Ile Ser Gln Ala Glu 145 . 150 . 155 . 160

Ala ·

<210> 50

<211> 39.

<212> PRT

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: protéine de fusion M-DP-TME1

<400> 50

Met Asp Pro Ile Ala Gly Ala His Trp Gly Val Leu Ala Gly Ile Ala 1 5 10 15

Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val Leu Val Val Leu Leu 20 25 30

Leu Phe Ala Gly Val Asp Ala 35

<210> 51

<211> 33

<212> PRT

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: protéine de fusion M-DP-TME2

<400> 51

Met Asp Pro Glu Tyr Val Val Leu Leu Phe Leu Leu Leu Ala Asp Ala

Arg Val Cys Ser Cys Leu Trp Met Met Leu Leu Ile Ser Gln Ala Glu 20 25 30

Ala



<210> 52

<211> 239

<212> PRT

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle: glutathion transférase (GST)

<400> 52

Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro 1 5 10 15

Thr Arg Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu 20 25 30

Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu 35 40 45

Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys
50 55 60

Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn 65 70 75 80

Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu 85 90 95

Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser 100 105 110

Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu 115 120 125

Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn 130 135 140

Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp 145 150 155 160

Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu 165 . 170 175

Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr 180 185 190

Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala 195 200 205

Thr Phe Gly Gly Asp His Pro Pro Lys Ser Asp Leu Ser Gly Gly 210 215 220

Gly Gly Leu Val Pro Arg Gly Ser Pro Gly Ile His Arg Asp 225 230 235

<210> 53

<211> 170

<212> PRT



<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:
 thiorédoxine (TrX)

<400> 53

Met Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser Phe Asp Thr Asp 1 5 10 15

Val Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe Trp Ala Glu Trp
20 25 30

Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp 35 40 45

Glu Tyr Gln Gly Lys Leu Thr Val Ala Lys Leu Asn Ile Asp Gln Asn 50 55 60

Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu 65 70 75 80

Leu Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val Gly Ala Leu Ser 85 90 95

Lys Gly Gln Leu Lys Glu Phe Leu Asp Ala Asn Leu Ala Gly Ser Gly
100 105 110

Ser Gly Ser Pro Lys Ser Asp Leu Ser Gly Gly Gly Gly Leu Val 115 120 125

Pro Arg Gly Ser Asp Pro Ile Ala Gly Ala His Trp Gly Val Leu Ala 130 135 140

Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val Leu Val 145 150 160

Val Leu Leu Phe Ala Gly Val Asp Ala 165 170